THE IMPACT OF CULTURE AND LANGUAGE ON THE USE OF THE INTERNET: EMPIRICAL ANALYSES OF BEHAVIOUR AND ATTITUDES

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät,
Humboldt-Universität zu Berlin

von

Anett Kralisch, M.A.
geboren am 31.01.1975 in Potsdam

Präsident der Humboldt-Universität zu Berlin (in Vertretung):
Prof. Dr. Hans Jürgen Prömel

Dekan der Wirtschaftswissenschaftlichen Fakultät:

Prof. Dr. Joachim Schwalbach

Gutachterinnen:

1. Prof. Dr. Bettina Berendt
2. Prof. Dr. Christa Womser-Hacker

Einreichung der Dissertation: 12.11. 2005
Tag des Kolloquiums: 16. 12. 2005

ABSTRACT

Internationalization and increasing border-crossing mobility are characteristic of the last few decades. It is a matter of fact that growing internationalization leads to cultures and languages in contact with each other more than ever before. The Internet can be considered as one of the most emblematic symbols of the internationalization process, yet one cannot ignore that barriers to accessing information on the Internet exist. These barriers are not only of a financial or technological nature, but are also founded in cultural and linguistic differences.

This thesis aimed to analyse the impact of culture and language on Internet use. Three main areas were investigated. These were:

(1) the impact of culture and language on preferences for information presentation and website structures, with a particular focus on search options,

(2) the impact of culture on preferences and the need for website content, and

(3) the impact of language within the context of the World Wide Web: language as a barrier to information access and as a determinant of website satisfaction.

We carried out nine studies where we opposed website users who used a website in their native languages (L1 users) to those who used the website in a non-native language (L2), where users with different cultural backgrounds were compared. In our analyses, we distinguished between a basic, stable concept of culture, as proposed by Hofstede, and a broadened concept of culture involving dynamic cultural aspects.

In order to test the 33 hypotheses, data was gathered by means of logfile analyses, online surveys, and laboratory studies. Depending on the methods employed, participants were either users of an international health website with a large variety in linguistic and cultural backgrounds, or local and international students at Humboldt-University Berlin and a major Malay university.

From our studies concerning culture, it was concluded that culturally determined thinking patterns, in terms of information need, space perception and time perception and structuring, clearly correlated with patterns of navigation behaviour and the use of search options. From these behavioural patterns we inferred preferences for website structures and information presentation that were grouped into two main types. However, the cultural groups also shared basic behavioural patterns, which raised the question of whether culture should be interpreted as a mediator of otherwise universal preferences.

In contrast, results concerning the impact of culture on the need for specific information were less conclusive, and indicated a less salient impact of culture in this area. It was found that the impact of the dynamic aspects of culture, in terms of domain knowledge and web experience, are stronger than that of stable aspects. Nevertheless, the stable cultural values of Power Distance and Individualism showed a significant impact on the website users' attitudes towards data disclosure.

From our studies concerning language, it can be concluded that language plays an important role for website access and website use. Results showed that significantly fewer L1 users than L2 users accessed a website. Only a part of the discrepancy could be explained by the fact that websites were linked to others in the same language more than to websites in other languages. It was furthermore found that among L2 users who accessed a website, those with low domain knowledge differ ed significantly in their behaviour and preferences from L1 users. This finding points out the important role of domain knowledge – potentially resulting from education – for website use and information access.

Finally, outcomes also suggested that a user's satisfaction with a L1 website is negatively correlated with their proficiency level in a potentially used non-native language; satisfaction with a L2 website was positively correlated with the respective L2 language. Results indicated furthermore that website satisfaction was negatively correlated with the perceived amount of native language information online.

Findings are hoped to provide guidance for cross-cultural and cross-linguistic adaptations of websites in particular, and other information systems in general. They also contribute to gaining insight into culturally and linguistically determined reasons for the digital divide as a first step to diminishing the existing divergences.

Keywords: Cross-cultural research, cross-linguistic research, multilinguality, digital divide, human-computer interaction, usability, information retrieval, user interfaces

ZUSAMMENFASSUNG

Internationalisierung und grenzüberschreitende Mobilität haben die letzten Jahrzehnte entscheidend geprägt. Die zunehmende Internationaliserung hat dabei zu einem sprach- und kulturübergreifenden Austausch in einem kaum gekannten Ausmaß geführt. Ein symbolisches Beispiel dieses Internationalisierungsprozesses ist das Internet. Weltweit existieren jedoch große Unterschiede in der Nutzung des Internets als Informationsquelle. Die Ursachen dafür sind nicht nur finanzieller oder technologischer Art, sondern sind auch im sprachlichen und kulturellen Bereich zu suchen.

Diese Arbeit widmete sich daher der Untersuchung des Einflusses von Kultur und Sprache auf die Nutzung des Internets. Drei Hauptgebiete wurden bearbeitet:

(1) Der Einfluss von Kultur und Sprache auf Nutzerpräferenzen bezüglich der Darstellung von Informationen und Websitestrukturen. Ein besonderer Fokus lag hierbei auf der Untersuchung der Nutzung von Suchoptionen.

(2) Der Einfluss von Kultur auf Nutzerpräferenzen bezüglich des Inhaltes von Websiteinformationen.

(3) Der Einfluss von Sprache im Kontext des World Wide Web: Sprache als Informationszugangsbarriere und als Faktor, der die Nutzerzufriedenheit beeinflusst.

In den neun Studien, die im Rahmen dieser Arbeit durchgeführt worden sind, wurden Websitenutzer, die eine Webite in ihrer Muttersprache nutzen (L1-Nutzer) denen gegenübergestellt, die die Website in einer Fremdsprache nutzen (L2-Nutzer), sowie Nutzer unterschiedlicher kultureller Herkunft verglichen. Die kulturellen Vergleiche basierten dabei auf zwei Konzepten des Begriffes *„Kultur"*; das Basiskonzept des Begriffes *„Kultur"* umfasst stabile kulturelle Aspekts im Sinne von Hofstede, das erweiterte Konzept dynamische Aspekte.

Daten aus Logfile-Analysen, Onlinebefragungen und experimentellen Untersuchungen bildeten die Auswertungsgrundlage für die Überprüfung der 33 Hypothesen. Die Teilnehmer der Untersuchungen waren vornehmlich internationale Nutzer einer mehrsprachigen Gesundheitswebsites sowie internationale und einheimische Studententen der Humboldt-Universität Berlin und einer großen Malaysisischen Universität.

Die Arbeit zeigt auf, dass kulturspezifische Denkmuster mit Navigationsmusters und Nutzung von Suchoptionen korrelieren. Auf der Grundlage dieser Verhaltensmuster wurden Präferenzen für Website-Strukturen und für Darstellungen von Informationen abgeleitet, die in zwei Hauptgruppen

aufgeteilt werden konnten. Die Ergebnisse zeigen aber auch, dass es grundlegende Verhaltensmuster gibt, die in allen kulturellen Gruppen wiederzufinden sind. Es stellt sich daher die Frage, ob Kultur als Mediator universeller Präferenzen aufgefasst werden sollte.

Der Einfluss von Kultur auf Nutzerpräferenzen bezüglich des Inhaltes von Websiteinformationen erwies sich als weniger eindeutig. Dieses Ergebnis weist daraufhin, dass die Rolle von Kultur in diesem Bereich weniger bedeutend ist. Der Einfluss dynamischer kultureller Aspekte (wie zum Beispiel themenspezifisches Wissen oder Interneterfahrung) war dabei jedoch stärker als der stabiler kultureller Aspekte. Nichtsdestotrotz konnte auch gezeigt werden, dass die in dieser Arbeiten untersuchten stabilen kulturellen Aspekte „Soziale Distanz“ und „Invidualismus“ signifikant die Einstellung der Nutzer gegenüber der Angabe persönlicher Daten im Internet beeinflussen.

Aus den Untersuchungen zum Einfluss von Sprache ging hervor, dass Sprache Websitezugriff und –nutzung beeinflusst. Die Daten zeigen, dass signifikant weniger L1-Nutzer als L2-Nutzer auf eine Website zugreifen. Nur ein Teil dieser Diskrepanz konnte damit erklärt werden, dass Websites gleicher Sprache stärker miteinander verlinkt sind als Websites unterschiedlicher Sprachen. Ein weiteres Ergebnis der Arbeit ist, dass L2 Nutzer mit geringem themenspezifischem Wissen sich in ihrem Verhalten und ihren Präferenzen signifikant von L1-Nutzern unterscheiden. Dieses Ergebnis weist auf die wichtige Rolle von themenspezischen Wissen – möglicherweise eine Folge des allgemeinen Bildungsniveaus – in bezug auf Website-Nutzung und Informationszugang hin.

Schließlich lassen die Ergebnisse auch darauf schließen, dass die Nutzerzufriedenheit mit einer L1-Website negativ mit den Sprachfähigkeiten in einer potentiellen L2-Sprache korreliert; die Nutzerzufriedenheimitt mit einer L2-Website hingegen scheint positiv mit den Sprachfähigkeiten in der betreffenden L2-Sprache korreliert zu sein. Die Resultate lassen außerdem vermuten, dass die Zufriedenheit mit einer Website negativ mit der wahrgenommenen Menge muttersprachlichen Angebots im Internet korreliert.

Ziel dieser Arbeit ist es, die sprachliche und kulturelle Anpassung von Websites und anderen Informationssystems durch neue Erkenntnisse zu unterstützen. Die Arbeit trägt außerdem dazu bei, mögliche Ursachen der vorhandenen digitalen Kluft aufzudecken, mit dem Ziel, die existierenden Unterschiede soweit wie möglich zu verringern.

Schlagwörter: Interkulturalität, Multilingualität, Mehrsprachige Websites, Digitale Kluft, Mensch-Computer-Interaktion, Usability, Benutzerfreundlichkeit, Informationssuche, Benutzeroberflächen

Meinen Eltern und Jason gewidmet.

Dedicated to my parents and Jason.

ACKNOWLEDGEMENTS

While working on my thesis I received invaluable guidance, advice and assistance from many people. I would hereby like to thank them for their generous support.

First of all, I would like to thank my supervisor Bettina Berendt. Her enthusiasm and commitment, and our numerous and endless discussions about the research's topics pushed me to work much harder than I would have without her. Her guidance was definitely invaluable.

I am grateful to my second supervisor, Christa Womser-Hacker, who encouraged my work a number of times with great interest. I also want to acknowledge her flexibility in supervising this thesis.

I am grateful to Oliver Günther, head of the Institute of Information Systems, who gave me the chance to conduct my work at the Institute. He provided important initial assistance and continuous logistical support, assuring a stable work environment for me to complete this dissertation.

Many thanks to Dieter Kattenbusch, a supervisor of my graduate studies in Romance languages for a number of years. He gave me the necessary confidence to start a Ph.D. thesis, and continued to provide assistance toward this goal even after I left his institute.

Furthermore, very important encouragement and assistance was provided by my office mates, my Institute colleagues, and numerous others. My office mates Claus Boyens and Max Teltzrow gave me extremely useful support and motivation throughout my thesis, especially in the initial phases. Over the last three years, they took responsibility for my good mood with their sunny dispositions, particularly during the difficult times that occur during every Ph.D. work.

I will never forget Matthias Fischmann's help. His unquestioned support at any moment, especially nightshifts full of programming work, saved important projects from impending deadlines. Furthermore, Steffan Baron and Gebhard Dettmar spent long hours in assisting with technical parts of data analyses, and were at the same time great teachers in these fields. I am grateful to the co-authors of my publications, without whose contributions writing these papers would not have been possible. Veit Köppen, Martin Eisend, Thomas Mandl, Alvin Yeo, and Nurfauza Jali provided inspiring content and supported my work with their technical input. Thanks also go to Rewadee

Anujapad and Marc Llop for their integral support in data collection, as well as to the translators, editors, and proofreaders of the multilingual questionnaires.

Throughout my doctoral studies I received a scholarship from the Graduate Program for young scholars (Nafög) as well as financial support from the faculty's Women's promotion program, which facilitated my work immensely.

Finally, I would like to send the deepest thanks to the persons to whom I want to dedicate this Ph.D. thesis.

Ich möchte meinen Eltern, Christiane und Burkhardt Kralisch, danken, die mich während meines ganzen Studiums von ganzem Herzen unterstützt haben.

Thank you, Jason, for the overwhelming and everlasting support you gave me during all the years of my Ph.D. thesis. Thank you for showing endless patience, for putting up with me in stressful moments, for keeping me well fed, and last but not least for your numerous hours and nights you spent proofreading.

Euch sei diese Arbeit gewidmet. This work is dedicated to you.

LIST OF CONTENTS

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

HCI	Human-Computer Interaction
IND	Individualism
IP	Internet Protocol
IS	Information Systems
IT	Information Technology
LTO	Long-term Orientation
L1	First Language (Native Language)
L2	Second Language (Non-Native Language)
MAS	Masculinity
OECD	Organisation for Economic Cooperation and Development
PD	Power Distance
STO	Short-term Orientation
UA	Uncertainty Avoidance
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WWW	World Wide Web
WUM	Web Utilization Miner

0. INTRODUCTION

0.1 ASPECTS OF CROSS-CULTURAL AND CROSS-LINGUISTIC RESEARCH WITHIN THE FIELD OF INFORMATION SYSTEMS

Internationalization and increasing border-crossing mobility are among the key characteristics of the last decades. It is a matter of fact that increasing internationalization leads to cultures and languages in contact with each other more than ever before in the history of human kind.

The Internet can be considered as the most emblematic symbol of the internationalization process: one of the key advantages of the World Wide Web is indeed its global accessibility. It permits information exchange and service offerings worldwide. It also allows an increasing number of people with different socio-economic, educational, linguistic, and cultural backgrounds to participate in this worldwide information and service market.

Yet one cannot ignore that barriers to accessing information on the Internet exist. Studies show how the distribution of the Internet differed and still significantly differs between countries, genders, or people of different educational backgrounds (Sims, et al., 2005). A number of reasons explaining these divergences have been identified, in particular those of a financial and technological nature.

Yet, financial and technological development can only partly explain Internet adoption and use. Given certain country-specific Internet usage patterns that did not fit the patterns of Internet adoption based on economic and technological development (e.g. Japan – see Hermeking, 2004), (see also Kralisch, 2003), researchers started to look into the roles of culture and language. The impacts of culture and language have been receiving increasing attention over the last few years, mirrored by the growing number of publications and conferences in this field. In view of the myriad of websites and (other) information systems that target an international audience, the question of what the consequences of different cultural and linguistic backgrounds are needs to be raised. How do language and culture affect the access of information, how do they influence the use of information systems, and what is their impact on the users' attitudes towards a website and their resulting satisfaction with it?

Besides these application-oriented questions, insight into the impact of culture and language may also contribute to the advancement of Information Systems from a theoretical point of view. Cross-cultural and cross-linguistic research helps to determine the extent to which (previous)

findings can be generalized. For instance, to which extent are privacy concerns universal? To which extent are they equally applicable to the U.S. and Europe? (see also Davison, et al., 2003). Cross-cultural and cross-linguistic research may also discover patterns where differences have been so far predominantly considered on the individual level. For example: Does culture lead to systematic differences between individuals in risk perception? (e.g. Weber, et al., 1998).

From a practical perspective, insight into these subjects is useful from a commercial point of view, since it enhances customer acquisition and services. The study of cultural and linguistic differences in this context can be interpreted as a form of marketing research that examines the adoption and user satisfaction of information systems. The study of language and culture is equally helpful from an ethical point of view: understanding the role of language and culture helps in making information and services available online for people with linguistic and cultural backgrounds that are so far underrepresented on the Internet. The goal of studying language and culture is to contribute to the decrease of the digital divide.

The research topic furthermore finds justification in the fact that adaptation of information systems to all cultural and linguistic groups is infeasible, for economic as well as technological reasons. Besides, such adaptations may represent important investment decisions, with ramifications that need decision support. Consequently it is important to determine aspects of language and culture that have a crucial impact on behavioural as well as attitudinal outcomes. It is important to discover which adaptations are necessary and which are less important; which user groups need to have separated information or service offers and which groups can be assembled into one cluster. In Section 2.4.1, for instance, we infer an assignement of culturally determined website preferences to two main groups from our results.

It is obvious that this thesis cannot give an extensive answer to the vast and complex role of language and culture in the use of information systems in general and websites in particular. Nevertheless, it provides another piece to the puzzle, and contributes towards obtaining a complete picture. This thesis focuses on the impact of culture and language on website access and website usage behaviour, valuation of features and (language) services, as well as attitudes towards data disclosure.

Culture is a rather recent subject in the context of Information Systems. As in other disciplines, analyses of culture within the field of Information Systems are challenged by a problematic conceptualization of the research object. What is culture? What are typical characteristics of a certain cultural group? How should a cultural group be defined? Is it possible that a person has

more in common with a foreign colleague who lives far away than with their neighbour, the postman, or the director of the local bank? Is this due to culture? Does cultural background have an effect on behaviour and attitudes? An analysis of the impact of culture requires a thorough theoretical discussion of the conceptual framework that precedes the empirical section in this thesis. As a first working definition, culture is defined as values, beliefs, and thinking patterns that are shared within a group. We elaborate more on this definition in section 1.3.3.1 in chapter 1. Since our empirical research is based on Hofstede's model, cultural groups are mostly defined on the national level. Sections 1.1.3.2, 1.1.3.3, 1.1.3.4, and 1.3.3.1 discuss this problem in more detail and suggest further units. Cross-cultural studies in Information Systems are additionally challenged by the fact that the cultural research paradigm that was originally developed in social sciences needs to comply with the requirements of a strongly application-oriented and quantitative research field. This needs to be taken into account when evaluating and interpreting cross-cultural research designs and results.

Language appears to be a less controversial subject. Yet, its role is multifaceted, and with regard to the Internet subject to contradictory impressions. It seems to be common sense that Internet users prefer websites in their native languages, especially among users with a lower proficiency level in English or other non-native languages. Nevertheless, it doesn't exclude that in certain situations non-native language skills are completely sufficient for understanding and navigating on websites. Especially during the beginning of the Internet, the use of English for offering information and services was widespread and widely accepted. Differences in users' language proficiency levels and their effects are examined when language is approached from a cognitive perspective. The cognitive approach to language regards mental processes, in particular in terms of information processing that is determined or affected by language. However, the use of language also triggers attitudinal patterns. Language attitudes are a classic research field in sociolinguistics. The attitudinal approach emphasizes affective responses towards the use of a particular language.

Since a comprehensive linguistic adaptation is not possible, the question as to which extent the use of (a) particular language(s) leads to noteworthy cognitive, attitudinal, and eventually behavioural differences regarding access of information and use of specific features is raised. Hence, examinations are needed to gain insight into divergences between speakers of different native languages and their attitudes towards native and non-native language offerings. As argued later on, the answer to this question is closely related to the extent to which language constitutes a barrier to the flow of information on the Internet. Given the decreasing impact of legal and logistical borders on the Internet, insight into the role of language with regard to this aspect seems to be of utmost

importance. Indeed, increasingly facilitated international trade and minimized communication borders will make more and more political and geographical borders disappear. It can be assumed that especially with regard to the Internet, linguistic borders are therefore going to play a much more important role in the future – as a market determinant from a commercial perspective or as a barrier to information access from a societal point of view.

To sum up, this thesis aims to contribute towards answering questions concerning the impact of language and culture on the use of the Internet and its services. Our investigations test the role of culture and language with respect to a number of behavioural and attitudinal aspects of website access, website use, and website evaluation. Analyses embrace several website features and services and potential implications for website design.

Studies are based on investigations of user behaviour on Ehealth-websites. Health websites represent an appropriate type of website since they encompass both societal and commercial goals: as part of their public information task, they aim to provide information to as many people as possible worldwide, regardless of their cultural and linguistic background. Moreover, they often aim in particular at people with an increased need for health education - usually the result of a lack of access to information. At the same time – due to increasing financial restrictions in the Health sector – remunerated services are becoming more and more mandatory for online health information providers. Customer services are therefore also evaluated under the aspects of their potential commercial benefit.

0.2 STRUCTURE OF THE THESIS

The Thesis is structured as follows: Chapter 1 presents the conceptual framework of the two research objects: language and culture. This includes a discussion of the methodological challenges of the analysis of culture, focussing in particular on the research field of Information Systems. Since our analyses are primarily based on Hofstede's data and cultural concept, advantages and drawbacks of his quantitative research paradigm are compared, and a broadened concept of culture is suggested. We propose a classification of how language can be analysed in the context of Information Systems: it distinguishes between a user's perspective and a provider's perspective, as well as between cognitive and attitudinal processes. Chapter 1 is concluded by a revision of the methods and constructs employed within our research.

This theoretical part is followed by the empirical part, which is subdivided into three areas. Each area is considered in a separate chapter.

Chapter 2 investigates the impact of language and culture with regard to users' navigation and information retrieval behaviour. We present four studies that relate to the impact of culture on navigational patterns and the use of search options, as well as the impact of language (and domain knowledge) on information seeking behaviour and – as a counterpart - preferences for information categorization. Language and culture are examined mainly under cognitive aspects.

Whereas the studies in chapter 2 emphasize cognitive processes and behavioural outcomes, the studies that are presented in chapter 3 focus on the impact on attitudinal variables. It is furthermore restricted to the analysis of culture. In four studies we examine the users' needs for information and their resulting risk perception and product valuation, as well as their attitudes towards disclosing personal data. The suggested broadened concept of culture is applied.

Chapter 4 complements results from chapter 2 that regard the impact of language. We examine the impact of language on hyperlink setting and hyperlink following behaviour, inferring the role of language as a barrier to information flow and access to information on the Internet. The second and third study of this chapter focus on attitudinal aspects, and investigate the users' satisfaction with a website depending on their foreign language proficiency skills and the perceived amount of native language information online. They complement chapter 3 since they add attitudinal aspects that regard language.

The thesis is concluded with a summarizing discussion of the results and existing limitations of our research, leading to an outlook on future research. At the beginning of the section a simplified overview of the studies conducted is provided, as well as their results.

Several parts of this thesis are based on previously published work. Essential parts of results regarding the impact of culture, language and domain knowledge – here discussed in chapter 1 - were published in (Kralisch and Berendt, 2004; Kralisch and Berendt, 2004; Kralisch and Berendt, 2005; Kralisch and Berendt, 2005). The impact of culture on navigation behaviour has been previously discussed and analysed in (Kralisch, et al., 2005). Results from study 4 – presenting linguistically (and culturally) determined preferences for information categorization – will be published in (Kralisch, et al., 2006). The role of culture with regard to risk perception and risk reduction behaviour as presented in study 5 (chapter 2) has been examined earlier in (Kralisch and Eisend, 2005). A previous version of analyses and results presented in study 8 (chapter 4) will be published in (Kralisch and Mandl, 2006). Finally, the conceptual framework of study 9 (chapter 4) is based on (Kralisch and Köppen, 2005).

Within the thesis we refer furthermore to (Kralisch, 2003; Kralisch, 2005; Kralisch and Mandl, 2005).

1. CHAPTER:

CONCEPTUAL AND METHODOLOGICAL FRAMEWORK

1.1 CULTURE

The role of cross-cultural studies in Information Systems research is many-fold. Academic literature ranges from the management of the distributed use of information systems in international organizations (Dafoulas and Macaulay, 2001; Yoo, et al., 1999) to adaptations of interface design for specific target cultures (e.g., Evers, et al., 2001). It also covers the role of cultural adaptation as a marketing tool used to identify competitive advantages. With growing awareness of cultural issues and its increasing integration into business and technology related areas, the concept of culture and its role in Information System and HCI[1] research are discussed more than ever. Applied[2] cross-cultural research such as in Information Systems is constantly challenged to comply with the balance between theoretical and academic suitability on the one hand, and its appropriateness for practical applications on the other. An attempt to provide clarification to this dilemma leads to two crucial questions that will be addressed within the scope of this chapter: *What is the goal of cross-cultural research in Information Systems?* and *How should culture be defined in the context of IS research?* Giving appropriate answers is particularly challenging since both questions mutually depend on each other and are subject to context-dependent interpretations, as argued in the next paragraphs. Both questions are mirrored in the methodological challenges cross-cultural IS research encounters when measuring culture as a construct. The following paragraphs aim to introduce current argumentation

---

[1] Information Systems: The entire infrastructure, organization, personnel, and components for the collection, processing, storage, transmission, display, dissemination, and disposition of information . **ATIS (2001): National Information Systems Security (INFOSEC) Glossary, 2005, August 3, http://www.atis.org/tg2k/_information_system.html**

HCI: Short for ***h****uman-****c****omputer* ***I****nteraction*, a discipline concerned with the study, design, construction and implementation of human-centric interactive computer systems (e.g.. http://isp.webopedia.com/TERM/H/HCI.html; last visited: August 11th 2005, see also subgroups the ACM classification system).

According to the ACM Classification System (1998) HCI (H.5., in particular H.5.2) is a subgroup of Information Systems (H.) Our research is best described by the term *HCI* but refers also to other subareas of Information Systems (e.g., Information Search and Retrieval - H.3.3; Human factors/Human Information Processing – H.1.2). Besides, some parts of this thesis could also be classified into top level areas other than Information Systems: for example, J.4. – Social and Behavioral Sciences). In the following, we the term *"Information Systems"* is used as the superordinate term for the research areas of this thesis.

[2] Applied research designates research conducted to acquire new knowledge with a specific application in view. It is opposed to basic research, although both kinds of research often overlap **Clarke, R. (2000): Appropriate Research Methods for Electronic Commerce, 2005, October 1, 2000, April 19, http://www.anu.edu.au/people/Roger.Clarke/EC/ResMeth.html**.

lines present in cross-cultural IS research. These involve the objectives of cross-cultural IS research, different perspectives of defining culture, and methodological issues discussed against the background of a persistent interrelationship between these three subjects (see also Bodley, 2000). Since this thesis' goal is the application of the construct of culture rather than its theoretical discussion, we limit the discussion to a number of selected argumentation lines. As a synopsis of our theoretical considerations, we present the definitions of culture that were adopted for our own analyses at the end of this chapter in the methodological section (see also Clarke, 2000).

1.1.1 Introduction to Cross-Cultural Research: a Brief Historic Overview with Particular Focus on Hofstede and Cultural Dimensions

"Interest in other cultures is probably as old as the exposure of human tribes to other tribes (Dahl, 2004). However, a first scientific discussion of the term *"culture"* regarding its content and consequences started only in the 19th century as part of anthropological science. In 1871, Edward Burnett Tylor published his book *"Primitive Culture"* (Tylor, 1871) which is considered the beginning of cultural anthropology (e.g., Dahl, 2004; Holden, 2002). The goal of these first studies of culture differed considerably from those of modern analyses and can only be understood if the historical context is taken into account. From a contemporary perspective, anthropological research was at that time "in service of imperial power" (Holden, 2002, p.21). Given this framework, the goal of these early studies consisted essentially in analysing culture as a form of developed civilization, opposed to barbarianism. Comparing these "…various stages of civilization among races known to the history…", early cultural anthropology also aimed to identify "…causes which have produced the phenomena of culture" (read: civilization.

The modern concept of cultural research as employed in many application areas such as Information Systems was developed no earlier than in the late 1950s, starting with the publication of Edward Halls' influential book *"The silent language"* (Hall, 1976; Hart, 1996). These new approaches differed from those of earlier cultural studies since they no longer perceived *culture* as a hierarchy generating term. Instead, *culture* was introduced as a term that describes various forms of civilizations at equal levels. Without prejudice, these new approaches embraced the first attempts to identify and classify typical patterns of different cultural groups. Cultural research became mainly comparative *cross*-cultural research (cultural relativism[3] – (e.g., Lévi-Strauss, 1966). At the same time, academic research aimed at overcoming cultural gaps by making people aware of these differences.

[3] "Cultural relativism affirms that one culture has no absolute criteria for judging the activities of another culture as "low" or "noble". Lévi-Strauss, C. (1966): The Savage Mind, University of Chicago Press, Chicago.

In particular, modelling culture as an iceberg that consists of a small visible protrusion (above the water's surface) and a bigger invisible mass (below the water surface) justified challenges in cross-cultural communication and, as a consequence, cultural education was used to overcome these. Accordingly, cultural literacy was expected to be the solution to escape the hidden constraints of the newly discovered, covert aspects of culture. Over the years, Hall was followed by authors such as Strodtbeck & Klockhohn (*"Variations on Value Orientationl"* 1961), Hofstede (1980 *"Culture's consequence"*, 1991 *"Software of the mind"*), Adler (*"International Dimensions of Organizational Behavior"* 1991), Schwartz (*"Universals in the Content and Structure of Value: Theoretical Advances and Empirical Tests in 20 Countries"* (1992), Trompenaars & Hampden-Turner (1993) *"Riding the waves of culture"*).

Extended summaries of cultural studies and their contribution towards insight into intercultural differences can be found in Hoeckling (1995), Hoft (1996), Darlington (1996), Holden (2002, p.21), and Dahl (2004). A rather exhaustive listing of existing approaches to cultures is published in Baumgartner (2003). Our work is governed to give a brief overview of the most important scholars and their contribution to the cross-cultural research field. The overview adapted from Baumgartner (2003) is depicted in the following table.

Table 1. Brief Overview of Cultural Studies

1951	Parsons	"The Social System"	5 pattern-variables: Affectivity-Neutrality, Universalism-Particularism, Diffuseness-Specificity, Ascription-Achievement, Instrumental-Expresssive
1959 1966 1976	Hall, E.T.	"The Silent Language" "The Hidden Dimension" "Beyond Culture"	Value Pairs: Communication: Context dependency Time: Polychronic vs. Monochronic Space: Intimate, Public, Social & Consultative Spaces
1961	Kluckhohn, F.R. & Strodtbeck, F.	„Variations in Value Orientations"	5 Dimensions: Relationship to Nature, Time, Character of Human Nature, Human Action, and Relationship to Others

1980 1991	Hofstede, G.	"Culture's consequences" "Culture and Organizations: Software of the Mind"	5 Dimensions: Power Distance, Collectivism, Femininity vs. Masculinity, Uncertainty Avoidance, Long-Term Orientation
1991	Adler, N.	"International Dimensions of Organizational Behavior"	6 Dimensions: The Nature of the Individual, The Relationship of People to Their World, Individualism vs. Collectivism, Doing vs. Being, Time Orientation, Space Orientation
1992	Schwartz	"Universals in the Content and Structure of Value: Theoretical Advances and Empirical Tests in 20 Countries"	Change from behavioural questions to behavioural values (How *important* are these values as "guiding principles of one's life"); Schwartz Value Inventory; 10 distinct value types
1992	Victor, D.A.	"International Business Communication"	LESCANT model: **L**anguage, **E**nvironment and Technology, **S**ocial Organization, **C**ontexting and Face-Saving, **A**uthority Conception, **N**onverbal Communication and Face-to-Face Negotiations, Conceptions of **T**ime
1993	Trompenaars, F.	"Riding the Waves of Culture: Understanding Cultural Diversity"	7 Dimensions: Universalism-Particularism, Individualism-Communitarism, Specific vs. Diffuse, Affective vs. Neutral, Achievement vs. Ascription, Sequential vs. Synchronic, Internal vs. External Control

The table reveals a growing use of so-called *"cultural dimensions"* in academic literature. In fact, the increasing interest among non-anthropological scientists studying culture has raised the demand for a new, quantifiable construct of culture over the years. Consequently, the need for extensively structured approaches to culture became more salient, thereby reducing the inherent complexity of the topic. With an increasingly businesslike orientation of cross-cultural research, cultural

characteristics were predominantly described in the form of dichotomies where a society's value orientation is ranked somewhere between the two extremes of the continuum – the cultural dimensions. Through the introduction of these dimensions, different cultures became, to a certain extent, comparable. Hall introduced the cultural dimensions of high context and low context describing different forms of communication (Hall, 1976). In an earlier work he had opposed monochronic and polychronic time orientations (Hall, 1959). Other authors added further dimensions describing people's relationship with time, nature, and other human beings[4]. Hofstede suggested five dimensions in total [5] (Hofstede, 1980; 1991; 2001), Trompenaars and Hampden-Turner (1993) seven. The use of dimensions therefore became a popular approach: not only did it provide a useful way means to characterize single cultures, but it also allowed academics to compare cultures among each other. The ten dimensions of Schwartz's value inventory (Schwartz, 1992) somewhat differ from the other authors' dimensions since they are based on *how important* certain values are to a society. With the introduction of the concept of cultural dimensions where cultural groups are described relatively to each other, the cultural research approach became a *cross*-cultural research paradigm consolidating the comparative nature of this research field.

The publication of Hofstede's *"Culture's Consequences"* marked an important milestone in the history of cross-cultural research. With his large study of IBM employees in 70 subsidiaries worldwide, the Dutch anthropologist was the first and so far only scholar to offer extensive empirical data that allowed an extensive quantification of culture along various dimensions. Hofstede's work provided new impulses in cross-cultural research where culture was increasingly approached from a quantitative point of view. By expressing cultural characteristics through numbers (*"cultural index scores"*), Hofstede made the concept of culture largely quantifiable and provided the necessary basis for conducting quantitative empirical research within the field of cultural research.

For instance, according to the Index Score for Individualism (IND) Chile is more collectivistic (IND=23) than Argentina (IND=46). Moreover, Argentina is twice as individualistic as Chile. Compared to other countries, in particular European and North-American, and given the range of 0 to 100[6] for Hofstede's cultural index scores, both south-american countries can be considered as

---

[4] According to Trompenaars and Hampden-Turner, a society's culture is manifested in its relationship with these three core issues

[5] Hofstede's original work included four dimensions. The fifth dimension – Long-term Orientation – was added later .

[6] The only exceptions are the Index Score for Long-term Orientation for China (118) and the Uncertainty Avoidance Index Score for Guatemala (101).

rather collectivistic (e.g., USA: IND=91) (see also Appendix A-1.2 for more index score data and description of cultural dimensions).

Statistical data has been previously published by other authors but was not sufficient for conducting extensive cross-cultural research (e.g., Hall, 1976). Despite growing criticism towards Hofstede's cultural paradigm over the years, his cultural index scores were employed in a myriad of cross-cultural studies, making Hofstede one of the most cited authors of this century (Social Science Citation Index). A big part of Hofstede's success can therefore be ascribed to the fact that cultural classifications through numbers constituted a reduction of culture's inherent complexity, making it available for quantitative research approaches. In addition, the quantification of culture through Hofstede's index scores also coincided with an existing dominance of quantitative research over qualitative research in human science studies[7] (see also Lamnek, 1988). Further views on the quantitative – qualitative debate are presented in section 1.3.1.2.

In 2002 Holden points out that after so many years of referring to Hofstede "there is a discernible feeling among scholars" to overcome Hofstede and reach for new approaches (Holden, 2002, p.11). Indeed, considering the large number of studies that are based on Hofstede, one easily gets the impression that Hofstede's approach has been overused, over-interpreted, and oversimplified. Nevertheless, we argue that not all criticism is justified nor applicable for all research areas.

In fact, Hofstede's work has provided an invaluable basis for quantitative cultural research. Yet, Hofstede's means of measuring culture is only appropriate if the implied conceptualiziation of culture is found to be suitable. As we shall see later, criticizing the application of Hofstede's models with empirical research does not relate to problems of measurement; it relates to his implied paradigm. However, as mentioned in section 1.1.3, the suitability of a definition and model of culture depends on the research objectives. In the following paragraphs we therefore focus our theoretical analysis of culture on the context of IS research, identifying research objectives in this field and examining the usefulness of Hofstede's model within this context.

Within this section it was illustrated how the concept of "culture" altered its meaning over time. The next section goes into more details of what the term "culture" represents within modern cross-cultural research. Up to date, scholars did not agree on a single meaning of the term *"culture"*. Accordingly a myriad of definitions exist.We presentation in the introduction of this thesis a first

[7] For an epistemological discussion of the usefulness of qualitative and quantitative research, please refer to section 1.3.1.

working definition in order to allow for a use of the term within our theoretical discussions. At the end of this chapter we will present a definition that is adapted to the context of IS research and to the purpose of research. In preparation of developing such an adapted definition, we elaborate in the next section on aspects of the concept of culture that are shared by the majority of definition. We also take into consideration aspects of the concept of

1.1.2 What is Culture? In Search for the Conceptual Core of Culture.

Despite the growing importance and popularity cultural research has gained over the years in anthropology, sociological and various other applied disciplines, scientific approaches towards culture appear genuinely difficult and particularly contested. Defining culture and measuring it have been the two major challenges to the academic community of cultural research.

One of the major challenges of cultural research is to define the content and limits of its research object. Conceptions of culture may indeed considerably differ between researchers and research projects, which is the reason why a discussion of the term is vital for every culture-related academic work. It was shown above how the meaning of the term *"culture"* has changed over the years. Until now a universally accepted definition does not yet exist. The lack of agreement in academic literature on defining the meaning of the term *culture* has been pointed out by a number of authors. In 1961, Kluckhohn and Strodtbeck counted 164 different definitions (Kluckhohn and Strodtbeck, 1969). It can be assumed that by today their number has increased by far. Nevertheless, we want to acknowledge that over the years a number of semantic cores have recurred in the works of the majority of authors. Contemporary anthropological views seem to have agreed on describing culture as descriptive, inclusive, and relativistic (Bodley, 2000). "*Descriptive*"[8] refers to a collection of various characteristics of the physical and social world that does not assign judging value, "*inclusive*"[9] "ascribes cultural significance to whatever […] has empirically demonstrable significance" and *"relativistic"*[10] indicates the comparative character (Johnson, 1998).

The most important definition nuclei as defined by Bodley (2000) are listed here below. All together they represent the semantic baseline of the term culture for our subsequent discussions of the role of culture in IS research.

---

[8] "descriptive" approach constrasts the "normative" approach where cultures are judged (usually as superior or inferior to another culture).

[9] "inclusive" vs. "exlusive" where only selected items and issues are regarded that would apply to all cultural studies.

[10] "relativistic" constrasts "absolute" where all cultures are compared with the same/universal norm/worldview.

Table 2. Semantic Nuclei of a Definition of Culture (Bodley 2000)

Topical:	Culture consists of everything on a list of topics, or categories, such as social organization, religion, or economy
Historical:	Culture is social heritage, or tradition, that is passed on to future generations
Behavioral:	Culture is shared, learned human behaviour, a way of life
Normative:	Culture is ideals, values, or rules for living
Functional:	Culture is the way humans solve problems of adapting to the environment or living together
Mental:	Culture is a complex of ideas, or learned habits, that inhibit impulses and distinguish people from animals
Structural:	Culture consists of patterned and interrelated ideas, symbols, or behaviours
Symbolic:	Culture is based on arbitrarily assigned meanings that are shared by a society

From these semantic cores can be inferred that the concept of culture cannot be applied to individuals. The *historical* and *behavioural* facets of culture, the fact that values and behaviour are transmitted from one generation to the next, and the fact that they are shared all emphasize that culture is specific, constitutive and distinctive for a *group* of people. A cultural group is accordingly defined as individuals who share a common culture, including common values and beliefs. The definition of what a group is can however differ as elaborated later in this section. In recent cross-cultural research the notion of a context-dependent cultural paradigm starts to prevail, taking the co-existence of different forms of cultural groups into account. Section 1.1.3.4 discusses these aspects.

Furthermore, the difficult role of "*behaviour"* as a characteristic of culture should be pointed out. *"Behaviour"* is in the area of psychology defined as "the aggregate of responses or reactions or movements made by an organism in any situation" (e.g. www.cogsci.princeton.edu/cgi-bin/webwn, last visited: November 1st 2005). In psychological research and in particular in cross-cultural psychological research, values are among the determinants of behavioural reactions. Kluckhohn (1951, p. 395) defined value as "a conception, explicit or implicit, distinctive of an individual or characteristic of a group, of the desirable which influences the selection from available modes, means and ends of action". Definition and conceptualization of *"values"* goes back to Allport

(1931). Rokeach (1979), Hofstede (1980; 1991; 2001) and Schwartz (1992) provided further important contributions. The question is raised of whether or not behaviour ought to be understood as a constitutive part of culture.

As shown above, Bodely incorporates *"behaviour"* into the characterization of a specific culture. With his view he differs from most authors who differentiate between the role of behaviour and that of value as constitutive parts of culture.

Hofstede, for example, symbolized the role of behaviour in the form of his *"cultural onion"* (Hofstede, 1991), depicting several layers of an onion's skin. According to Hofstede, values – as the heart of the onion - form the core of culture. They are however implicit, and can "only be inferred from the way people act under various circumstances" (Hofstede, 1991, p.8). Values are thus the deepest manifestations of culture (Hofstede, 1991, p.7), whereas behaviour – as the outer layers of the onion – is more superficial and results from these values. Nevertheless, the role of behaviour as a constuitive part of culture is not clear in Hofstede's cultural model. In fact, Hofstede uses the term *"practices"* which subsumes the three outer layers of the cultural onion: rituals, heroes, and symbols (Hofstede, 1991, p.8). It seems that the term *"practices"*, in particular the layer of *"rituals"* relates exclusively to behaviour that explicitly carries a cultural meaning (e.g. forms of greeting). Based on this interpretation, it can be inferred that behaviour that is determined by cultural values but without explicit cultural meaning is, in Hofstede's cultural model, not part of culture. Hofstede's model bears resemblance to the iceberg model. In the iceberg model the visible layer encompasses mainly behaviours (with an explicit cultural meaning), whereas the invisible layer includes for the most part implicit cultural values.

Holden very explicitly classifies behaviour as a consequence of culture and consequently excludes behaviour from being a part of culture (Holden, 2002).

In line with Hofstede's and Holden's perceptions, our own research is grounded in the assumption that (1) behaviour that does not carry an explicit cultural meaning is not part of culture but (2) a consequence of cultural values. Thus, measuring behavioural patterns reflects the underlying cultural values and therefore the core of each culture. In the following the term "behaviour" refers to this type of behaviour that does not carry an explicit cultural meaning.

As we shall see in the following section, theoretical models do not directly link behaviour to cultural values. Behaviour is rather the consequence of attitudes and/or mental models that are based on the cultural values individuals hold.

1.1.3 Objectives of Cross-cultural Research in Information Systems

As mentioned above, the appropriateness of definitions and methodological approaches is also dependent on the objectives of the conducted research. After having developed the theoretical basics of a concept of culture, the goal consists in specifiying and in applying these basic concepts to the context of IS research. An elaboration of the objectives of cross-cultural research within the context of information systems is therefore indispensable. It is the subject of this section.

Hofstede conducted his research mainly from the point of view of organizational management. In this area, identifying cross-cultural differences constitutes a contribution to facilitate the management of huge international organizations. Publications in cross-cultural organizational (e.g., Adler, 1991; Hofstede, 1991; Holden, 2002; Schein, 1985) and more specific in knowledge management (e.g., Holden, 2002) focus on the benefit that can (or should) be drawn from the interaction of different cultures. Culture is considered a solution rather than an obstacle for universal business processes. Interestingly enough, cross-cultural organizational and knowledge management are the research areas with the strongest attempts to overcome Hofstede's paradigm (see also 1.1.3). In this area insight is often gained from rather qualitative research, including theoretical considerations and case studies (see for example Holden, 2002). Qualitative research allows for the consideration of a broad spectrum of cultural aspects as well as the evaluation of finer details of single cases.

Hofstede's cultural index scores were also increasingly employed over the years in various forms of *"market research"* (e.g. consumer behaviour - De Mooij, 2003) and psychological research (e.g. non-verbal behaviour, cultural behaviour - Berry, et al., 1992; Ciborowski, 1979; Hofstede and McCrae, 2004; Matsumoto, 2000). In contrast to organizational and knowledge management, research efforts in market research and psychological research (as an auxiliary research area) focus mainly on assessing differences between cultures, for instance with regard to technology acceptance and use. As such, cross-cultural research helps to adapt international products to the users' needs and preferences. Cross-cultural research applied to this area is often of a predictive nature: it aims to infer insight about human behaviour from cultural variables and predict future reactions. From a business point of view, cross-cultural research increases competitive advantage through adaptation to these differences rather than through managing the differences. Due to its close connection to psychological research, empirical research is often of a quantitative nature, trying to find correlations or even cause-effect-relationships between culture and consumer behaviour. The quantitative approach, however, restricts the notion of culture to a simplified concept.

Identifying culturally determined local needs and preferences for the goal of product adaptation is also the main purpose of cross-cultural research in Information Systems. *"Preference"* is a concept mostly used in social sciences. "It assumes a real or imagined choice between the alternatives and the possibility of rank ordering of these alternatives" (http://en.wikipedia.org/wiki/Preference - last visited: November 2nd 2005). A high ranking expresses a preference and thus a predisposition in favour of something (http://wordnet.princeton.edu/perl/webwn - last visited: November 2nd 2005). *"Preferences"* should be distinguished from (cultural) *"values"*. As defined by Hofstede, "values are broad tendencies to prefer certain states of affairs over others." Values are the core of culture in accordance with Hofstede's "cultural onion" (Hofstede, 1991, p.8-9). Whereas *"values"* refer to *general* predispositions to favour something, apply to the entire social and physical environment, and are culturally determined, the term *"preference"* relates to very concrete things and might be determined through other variables than culture.

The term *"need"* is used in those cases where a lack of something is perceived by an individual as such. Needs are thus a matter of perception and therefore of subjective evaluations. In contrast to preference, *"need"* is of less comparative nature, since it can be assessed for a single product without contrasting it to others. However, degrees of *"need"* can be compared. *"Need"* increases with higher preference.

Sociological studies within IS research are similar to market research studies, and their vast majority aim to determine the acceptance of technologies and services within a defined market area. Cross-cultural research in Information Systems is therefore an application-oriented research field. However, it should be noted in the case of cross-cultural research that the use of the term "*market research"* embodies commercial as well as societal goals: it includes companies' analyses of consumer preferences for marketing products but also aims to enhance information systems employed in national health-systems, or to identify the needs of underrepresented groups. The crucial role of insight obtained from cross-cultural IS research in the effectiveness and efficiency of IT deployment is pointed out by Evaristo and McCubbrey (1995). Slay and Burke (2001) emphasize safety and security issues of culturally adapted information systems. In a similar manner as cross-cultural market research, cross-cultural research in Information Systems is dominated by quantitative approaches and simplified concepts of culture, in particular if the research serves application-oriented purposes.

Culturally determined technology acceptance and use depend on a variety of culture-related factors. Culture's most important impact consists in determining the *"objective usability"* in terms of cognitive

load, as well as the *"subjective usability"* in terms of user attitudes towards the system (see also Ford, et al., 2005). *"Objective usability"* and *"subjective usability"* represent the two main approaches in our studies. Both terms are aspects of the more general notion of *"usability"*. *"Usability"* is defined by the World Wide Web Consortium (W3C) as the *„effectiveness"*, *„efficiency"*, and *„satisfaction"* with which specified users achieve goals in particular environments (ISO 9241). *"Effectiveness"* is "the accuracy and completeness with which users can achieve specified goals in particular environments"; *"efficiency"* relates to "the resources expended in relation to the accuracy and completeness of goals achieved". *"Effectiveness"* and *"efficiency"* are the elements of usability that are part of *"objective usability"*. In contrast, *"satisfaction"* can be classified into the subgroup of *"subjective usability"*. *"Satisfaction"* stands for "the comfort and acceptability of the work system to its users and other people affected by its use".

Within cross-cultural research, *"objective usability"* relates to how culturally determined mental models (Honold, 2000) influence the perception of the complexity of a problem, the capability to maintain concentration, and the level of anxiety. Mental models – based on respective cultural values - constitute a representation of the real world that people use to understand specific phenomena. According to Norman (1983) mental models evolve when interacting with the environment and with other people. Members of the same cultural group are therefore supposed to share similar mental models. Hofstede's metaphor of culture as *"software of the mind"* illustrates once more how culture affects, to a large extent, the formation of mental models. Honold (2000, p. 5) points out that the interpretation of new information is based on these mental models. The level of correspondence between existing mental models and mental models implied in the information system influences the perception of concrete data, the level of concentration during reasoning and problem solving, and may – as mentioned above - even reduce anxiety[11] (Marcus, 2004). As a result the level of cognitive load and cognitive resources is affected. In addition, the user is only able to assign a useful meaning to the used information system if the employed symbols, icons, languages, etc. can be easily interpreted and correctly comprehended by them (e.g., Bourges-Waldegg and Scrivener, 1998).

In our research, we analysed the cognitive role of culture with regard to typical patterns in navigational structures (Kralisch, et al., 2005), with regard to search option use (Kralisch and Berendt, 2004; Kralisch and Berendt, 2004; Kralisch and Berendt, 2005), and with regard to aspects of information categorization (e.g. Kralisch, et al., 2006). Results are presented in chapter 2.

[11] The example of anxiety shows that sometimes the clear distinction between cognitive and subjective acceptance can be blurred.

A more subjective approach to culturally determined technology acceptance is derived from cultural values, beliefs, and attitudes that are shared within the cultural groups. (Jagne, et al., 2004; Russo and Boor, 1993). In contrast to *"objective usability", "subjective usability"* is a belief-based concept that is founded on users' perceptions rather than on actual self or system performance (Ford, et al., 2005). The formation of attitudes towards a system is based on these beliefs. "An attitude represents a person's general feeling of favorableness or unfavorableness toward some stimulus object. […] As a person forms beliefs about an object, he automatically and simulataneously acquires an attitude toward that object." Beliefs and consequently attitudes are formed through cultural values (e.g. Spencer-Oatey, 2000). Here again, members of the same cultural group are therefore likely to share the same beliefs and attitudes towards a system which results in similar usability (see also Carey, et al., 1998).

Chapter 3 presents results from our studies that regard these subjective aspects. They refer to the cost-benefit trade-off of visiting a website, identifying cultural differences in risk perception (Kralisch and Eisend, 2005), product valuation, and disclosure of personal data. Empirical investigations are however restricted to the assessment of culture related differences in attitudes.

1.1.4 Conceptual and Methodological Challenges of Cross-cultural IS Research

Cross-cultural research is challenged by a number of difficulties which are mostly of conceptual nature. With regard to some aspects, the application orientation of cross-cultural IS research and implied general tendency towards quantification reinforces these challenges.

Intercultural research in Information Systems is a relatively new research area that has gained increasing importance over the last few years (Evaristo, 2004). In 2002, an article by Karahanna et al. (2002) still described cross-cultural information systems research as relatively undeveloped. Considering the large number of unresolved research questions, even today much more research needs to be done, as Karahanna and her colleagues point out (2005). According to the same authors (Karahanna, et al., 2002) the lack of sufficient research in this field can mainly be explained by "methodological and resource difficulties inherent in cross-cultural research as well as the long time horizon required to complete and conduct these types of studies". Methodological issues of measuring culture and its impact on use of information systems have been also identified by other authors as a major challenge in cross-cultural IS research. (Troman, et al., 2005; Van De Vijver, 2000). In fact, it appears that methodological issues have become the central problem of cross-cultural research. In the context of IS research measuring culture refers essentially to two major challenges: first, the challenge of correctly measuring culture itself, i.e. classify cultural groups, and

second, the challenge of correctly measuring the *effects* of culture such as the need or preference for a particular system. For instance, Mullen (1995) requires for valid cross-cultural comparisons cross-cultural equivalence of measurement scales to imply that subjects across cultures respond to measurement scales in the same way. Further examples are given in (Karahanna, et al., 2002).

Whereas Karahanna and her colleagues focus exclusively on measuring problems that are inherent to comparative research in their article (e.g., bias, achieving cross-cultural equivalence as prerequisite), we emphasize unresolved ambiguities in measuring culture itself as the source of existing methodological issues in cross-cultural research. The reason for these ambiguities lies in the aforementioned interrelationship of methodological approaches, conceptualizations, and research objectives. In particular, the concept of culture (especially its semantic boundaries) as well as the objectives of cross-cultural research and the context in which it is conducted, are highly interdependent with methodological issues. A discussion of methodological difficulties can consequently not occur without examining these particular aspects of cross-cultural research. The following section aims therefore to discuss an adaptation of the concept of culture to the needs of IS research and further outline the interdependency between cultural paradigm, research objectives and fundamental methodological issues. At the end of this chapter we present concepts of cultures that were employed in our studies. Methodological difficulties that concern a specific research design are discussed in their respective sections.

As described earlier, empirical data from Hofstede has been the first and so far the only comprehensive data source that allows for the use of the concept of culture in quantitative research. It was therefore also the main data source for our analysis. Nevertheless, referring to Hofstede's data also involves relying on his cultural paradigm, which is largely contested within the academic community.

The discussion of the conceptualization of culture and methodological approaches is therefore focussed on Hofstede's model and its appropriateness for IS research objectives. Four key aspects of his model are discussed: causality in cross-cultural research, stability of culture, definition of cultural units and homogeneity/heterogeneity of cultural units.

1.1.4.1 Causality in Cross-cultural Research

One of the major challenges in defining the semantic boundaries of the term culture is the distinction between the causes of culture, culture itself, and the consequences of culture. At a first glance such a distinction seems simple and evident. In fact, numerous articles have provided

evidence for how cultural values are reflected in behavioural patterns of daily life. For instance, a myriad of correlations between the cultural index scores and socio-economic variables were found (e.g. consumption of mineral water, cigarettes, soft drinks - Masculinity, media coverage - Individualism, see De Mooij, 2003).

However, earlier in this chapter we mentioned the question of whether behaviour should be considered a constitutive part of culture or rather a consequence of culture. This debate can be interpreted as a first indicator of the unclear causality in cultural research. Yet, whereas this question might have primarily impacted on a theoretical level, the problem of the direction of the causality is much more concrete. Correlations as described above do not necessarily provide insight about the direction of causality. Therefore culture could be the reason for a specific degree of productivity (e.g. correlation between Individualism and GDP - Hofstede, 1980; Hofstede, 1991). On the other hand, GDP could explain why certain values and beliefs have developed within a group. A powerful example is given by the positive correlation between a society's degree of Individualism and its wealth (Hofstede, 1991). The correspondence between Individualism and wealth might lead to the conclusion that individualistic societal structures are more appropriate for a higher economic output. However, Hofstede himself claims that the correlation exists because a higher degree of Individualism results from higher prosperity rather than the oppositely directed causality. Depending on one's world view, Hofstede's interpretation might seem intuitive or counterintuitive, either right or wrong. Nonetheless, empirical scientific insight is limited to the simple identification of correlations without being able to draw any conclusions about the direction of causality. Prosperity therefore represents a good example of variables that cannot be clearly assigned as a consequence or a cause of a society's culture. Findings in cross-cultural research are therefore mostly limited to the identification of correlation where the causality can be interpreted in both directions.

Section 3.3.3 presents research results regarding culture-dependent privacy concerns, providing more specific details about the causality dilemma.

1.1.4.2 (In)Stability in Hofstede's Paradigm: A Broadened Concept of Culture

The question as to which degree environmental settings or socio-economic variables are a constitutive part of a cultural group's idiosyncrasies can be raised, and should therefore be incorporated as explanatory variables into cross-cultural analyses.

However, the integration of these kinds of variables encounters two problems. First, in Hofstede's model, socio-economic variables are not part of culture. Their incorporation would consequently require a broadening of Hofstede's concept and quantitative material. Second, socio-economic variables underlie fast-paced changes. Yet Hofstede's model presumes culture as a stable construct which appears to be in contrast with the nature of socio-economic variables. Nevertheless, we argue that stability in Hofstede's paradigm is not absolute: the impact of socio-economic variables – even if identical for each cultural group - makes the impact of cultural values on group characteristic *behaviour* relative. A broadening of Hofstede's concept should therefore be considered.

Relativity of Stability. As shown by Hofstede's recent work, changing socio-economic settings, in particular growing prosperity, leads to changes in people's behaviour such as saving behaviour (Webley and Nyhus, 1999). Cross-cultural market research found indicators of converging markets as well as diverging consumer markets (e.g. cars -De Mooij, 2003), although other studies seem to provide evidence for the strong impact of culture on consumer behaviour. At a first glance, these findings contradict Hofstede's assumption of the stability of cultural values. A changing of values as a result of changing incomes had indeed been predicted by economists (De Mooij, 2003).

However, one needs to distinguish between the change of cultural *values* and that of *behaviour*. Changing cultural values are a necessary precondition for changing consumer behaviour solely if cultural values are the only determinants of consumer behaviour. Yet there is sufficient empirical evidence that other variables such as the GNP per capita also influence consumption characteristics. These other variables are usually socio-economic variables, and are not part of a culture's core characteristics (which are usually defined by values) but still have an impact on consumption patterns that are characteristic for a group. This is particularly true if behaviour is measured on an aggregated national level, which is usually the case for these kinds of studies. Changes in consumer behaviour over time are thus more likely to be caused by changes in a country's socio-economic settings (or other variables) than by changes in cultural values. Referring to Hofstede's recent studies and replications of his IBM study (Hofstede, 2001), De Mooij (2003) points out that "cultural values are stable and with converging incomes they become manifest". Consequently, the "predictive/explanatory power of GNP/capita declines with converging national wealth" (De Mooij, 2003). The fact that with regard to certain products or to certain countries, consumer behaviour converges despite growing wealth finds its explanation in ceiling effects. For old products (e.g. newspaper, radio) ceilings were reached long ago and cultural differences have disappeared. In contrast, for new products (e.g. computers) the ceiling has not been reached yet (De Mooij, 2003) and consumer behaviour is largely influenced by socio-economic variables.

Similarly, car purchases in developing countries strongly are influenced by economic possibilities, whereas in richer countries culturally determined divergences in purchase patterns can be found (De Mooij, 2003).

Hence, worldwide growing prosperity leads to altered behaviour within a cultural group. The intensity of changes in behaviour depends on the strength of changes in the socio-economic settings of that group. There is however no indication that changes in socio-economic settings lead to changes in cultural values. Moreover, differences in cultural values become manifest with converging socio-economic settings. Yet, since socio-economic settings also determine behaviour, it is possible that changes in behaviour apply to each cultural group. Culturally determined *behaviour* (in constrast to values) is therefore not absolute, but should be primordially understood as a comparative measure: if fast walking at a speed of 5km/h was found to be a characteristic of highly individualistic countries (Levine and Norenzayan, 1999), the speed that is perceived as fast today could be characteristic of low individualistic/collectivistic countries in a couple of years. The balance between typical individualistic and collectivistic behaviour would be still relevant if people from highly individualistic societies walk faster in the future than they do today.

The discussion above also outlines the importance of the question of whether cultures in cross-cultural market research should be restricted to cultural values, for example as defined by Hofstede (1980; 1991; 2001).

Longterm vs. Shortterm Views of Stability. Socio-economic settings are fast-paced changing variables. Their impact on the single cultural groups converges only over time. As a result the impact of cultural values on behaviour might be obscured by the impact of socio-economic variables. The ranking order within the Hofstedian cultural system can therefore temporarily be out of balance. This would for example be the case if a high uncertainty avoidant culture experiences an outstanding economic growth in contrast to other cultural groups, leading to unusually high investment rates (for correlations between Uncertainty Avoidance and investment behaviour see Noorderhaven, et al., 1976-2000) that exceed those of cultural groups with lower Uncertainty Avoidance. During those periods, cultural values as determined by Hofstede's cultural index score provided little explanation for group specific behaviour. In fact, cross-cultural research often suffers from cultures' low contribution in predictive models. As Srite et al. (2003) remark, the variance in predictive models in classic cross-cultural research has been relatively low. These findings might be a first indicator that factors other than those implied in Hofstede's model, but that still are characteristic for a cultural group, also need to be taken into account.

Broadening Hofstede's Concept. Finally, mapping culture into bipolarized structures such as into Hofstede's cultural index scores, can never reflect culture in its comprehensive complexity. Using Hofstede's cultural index scores can therefore not assure that all shared values are reflected in these index scores. So far, there is no evidence that all important values and beliefs are assessed by these five dimensions[12] (see alsoMcSweeney, 2002).

Taking antecedents of culture such as socio-economic settings into account may also represent a useful way to provide further explanations for differences in behaviour, preferences, and needs that are specific to a cultural group, but that cannot be ascribed to one of the existing cultural dimensions. Besides, when measuring culture it is not always easy to distinguish the effects of values and attitudes that were acquired over generations from those that had emerged in the recent past, such as a country's economic situation (see alsoWeber, et al., 1998).

Depending on the study conducted, taking into account current socio-economic variables as a cultural specific trait can be particularly beneficial to the research goal. In line with Röse's incorporation of environmental conditions into the concept of culture (Röse, 2005), we argue that socio-economic settings should be classified as frame variables too. Röse distinguishes between three subgroups of cultural variables: direct, indirect, and frame variables. Frame variables subsume aspects that exert an influence on product use and acceptance, without having a direct impact on the interaction with the system (such as language, dialogue design). Examples are environmental factors, technological standards, education systems, or political systems[13]. Such a broadened approach to culture bridges one of the gaps of Hofstede's paradigm, since - according to his critics - behaviours and values are not wholly determined by cultural background (Williamson, 2002). We argue that such a broadened concept is particularly useful in cross-cultural IS research if the analysis of *current* conditions is the research objective.

In contrast, if a long-term analysis of the effect of culture is targeted, the impact of fast changing socio-economic settings should become less important as a driving force of culturally determined behaviour and preferences. However, at the same time the comparison of numerous results collected over time within the same cultural group but under changed socio-economic situations helps to detect the impact of culture as defined by Hofstede (i.e. values and beliefs).

---

[12] The number of four / five dimensions was determined through factor analysis that was carried out on the data set obtained from the IBM survey. Nevertheless, this outcome is mainly determined through the questions that were addressed in the questionnaire. Hofstede himself justifies the number of five dimensions with the human capacity to distinguish information.

[13] However, the author does not address the problem of measuring these variables.

Details about the integration of socio-economic variables in our study of culture are presented in the methodological section at the end of this chapter (see section 1.3.3.1). Within our research, such a broadened approach was applied in market-related analyses where the impact of culture on service evaluation and the need for service were evaluated (section 4.3).

1.1.4.3 Are National Units Appropriate for Measuring Culture in IS? Market-driven Cross-cultural Research Approaches

Cultural research involves aggregating individuals into cultural groups, i.e. to a group of individuals sharing similar values, beliefs, needs, and thinking patterns. It raises the question of how units of cultural groups ought to be defined. Optimal group segmentation should be characterized through homogeneous values and thinking patterns shared within each group, which are of a distinctive nature and lead to rather heterogeneous cultural characteristics between the groups.

Hofstede's model is based on cultural units that are defined on the national level. It is one of the most criticized points in Hofstede's paradigm, on the grounds that national borders represent artificial cultural frames rather than mirroring reality (Wright, 2000). The lack of correspondence between national and cultural borders is indeed a frequent counterargument to Hofstede's model. Yet, correspondence reflects only one possible motivation for defining cultural units by national boundaries. There are in fact other driving forces that make – as we argue – national boundaries more appropriate for applied cross-cultural analyses.

Hofstede himself declines this criticism by vindicating the coincidence between national and cultural borders. According to him, differences in values are usually more salient *between* countries than *within* countries. Research results provide evidence against and in favour of Hofstede's statement. Considering the countless possibilities to approach and evaluate this issue, it almost seems however that the truth in Hofstede's statement depends on the researcher's point of view.

Arguments in favour of a national culture concept are also provided by our suggested broadened approach to culture, involving socio-economic variables. The underlying reason is that socio-economic variables are often specific to a country. National legislation and policy leads to the fact that countries are often rather homogeneous within their borders and rather diverse from other countries with regard to socio-economic settings. Hofstede's approach of considering nations as cultural units is consequently justified in those cases, where socio-economic settings are considered to be an important part of the cross-cultural study.

This national approach finds a further justification in the fact that cross-cultural research often shares a number of common characteristics with classic market research. As Bourges-Waldegg et al. (1998) point out, defining culture has, in this case, more to do with defining market areas than defining real user groups. In these cases the impact of legal or logistic constraints (e.g. language laws, taxes, etc.) dominates market analyses. In the second step, cross-cultural research helps identify the characteristics of the predefined markets. Culture, when measured on the national level, can fit business investment schemes perfectly, and make it consequently appropriate for application-oriented cross-cultural IS research.

Finally, it should be noted that a definition of cultural groups based on nations loses its justification as soon as culture-determining variables are no longer specific and limited to single nations. So far, national legislation and policies, as well as trade borders and logistical limitations contribute to strengthen the perception of a nation as a cultural group. This is true with regard to the members of the cultural group themselves as well as with regard to those who offer culturally adapted products. However, increasingly facilitated international trade and minimized communication borders will make political and geographical borders disappear more and more. It can be assumed that especially with regard to the Internet, linguistic borders are going to play a much more important role as a market determinant in the future.

1.1.4.4 But each Individual is Different: Heterogeneity in National Cultural Groups

The lack of uniformity within cultural groups (however they are defined) and the use of stereotypes are further, often criticized points of Hofstede's model (e.g., Bourges-Waldegg and Scrivener, 1998; Fitzgerald, 2004; Jagne, et al., 2004; Light, 2003; Williamson, 2002).

It is widely accepted in academic literature (e.g., Hofstede, 1991, p.10) that each individual is at the same time part of several cultural groups defined at different levels. This means that cultural units here are defined along different categories: e.g. gender as one level vs. occupational knowledge as another level vs. nationality, etc. Accordingly, since the individuals of a national cultural group are members of different groups at the various levels they differ with regard to certain characteristics despite belonging to the same national group.

An answer to the question of which levels to emphasize in an investigation depends again highly on the research's objectives: Previous market analyses might for instance frame how to define the target group. Furthermore, it is the researcher's responsibility and expertise to estimate which variables should be incorporated into the analysis.

Dynamic approaches address the problem of a multiple group membership by introducing a context-sensitive approach (Briley, et al., 2002). They suggest that depending on the context, an individual's membership in either one or the other various cultural categories (e.g. job, gender, etc.) comes to the foreground. Through these dynamics "specific pieces of cultural knowledge […] become operative in guiding the construction of meaning from a stimulus (e.g., Hong, et al., 2000; Wong and Hong, 2005). Classic approaches that are based on context-dependent knowledge or value activation are the "Knowledge Activation Theory" (Higgins, 1996) or Tajfel's "Social Identity Theory". A key idea of the "Knowledge Activation Theory" is that "possessing a particular cultural construct does not entail relying on it continuously, and that predictable factors determine whether a construct will become operative" (Briley, et al., 2002). Tajfel's "Social Identity Theory" refers to the psychological process of identification and its motivational basis (Deaux and Martin, 2003), i.e. each individual has several "personal selves" that are usually triggered by the social context.

Finally, we argue that the lack of homogeneity within a national cultural group is not a shortcoming specific to Hofstede's approach. It is rather characteristic to any market research approach where cross-cultural research in IS can be considered (see above). The lack of uniformity within cultural groups is therefore hardly due to a poor group assignment, but rather suits market purposes. Aside from the impracticability of collecting data about each group's members, information about the values, attitudes, and behaviour of every single consumer is only useful if a product can be easily customized. If customization is not possible, market adaptation usually aims at satisfying the needs of the highest number of consumers possible, or focuses on a specific niche, but can never guarantee to respect everybody's requirements. Hence, Hofstede's cultural index scores can be understood as a description of the likelihood[14] of encountering people with certain characteristics. A description of country A as having low Uncertainty Avoidance is therefore not meant to disregard people with a rather highly uncertainty avoidant way of living and working, it just declares that people with a low uncertainty avoidant cultural background are more likely to be found within that country.

1.1.4.5 Unresolved Limitations: Each Model is only a Model

Despite the research communities' aspirations of developing alternative concepts, no viable alternative to Hofstede's data has been suggested. With the current state-of the art in methodology

[14] Hofstede's calculation of index scores is based on percentages of one particular answer per group as well as on mean scores. The term "likelihood" would be only exact with regard to percentages, whereas the mean scores express an average but do not necessarily provide information about probabilities.

for cross-cultural research, it appears that Hofstede's approach cannot be overcome if quantitative research and application orientation is targeted.

In section 1.1.3.2 we suggested broadening his concept of culture by taking current socio-economic variables and/or the individual's knowledge into account. In the methodological section at the end of this chapter, we present argumentations why part of our research was deemed justified for the national level and why for other parts modified approaches were more suitable. We furthermore provided justifications for defining cultural units on a national level. Yet Hofstede's data aggregation to the national level seems to be appropriate for IS research only as long as markets are defined by legislation and national borders. Hofstede's index scores become less useful if national borders disappear and other factors such as language[15] determine user groups to a much stronger extent, and therefore markets. Unfortunately, data aggregation to these levels cannot be deduced from Hofstede's cultural index scores.

Cross-cultural research based on Hofstede's paradigm also suffers from the fact that he is barely able to explain the short-term effects of cross-cultural interaction as it can be observed with increasing globalization (e.g. cultural convergences - Bryan, et al., 1994; Light, 2003; Norton, 2001) – or changes in cultural values for some cultures due to continuous pressure from other cultures (e.g., De Souza and Dejean, 1999). The integration of dynamic approaches into Hofstede's concept (see 1.3.3.1) allows us to consider the membership of an individual to different cultural groups as long as these groups concern different levels of culture (e.g. national and socio-economic affiliation) Yet Hofstede's model becomes problematic in cases where an individual belongs to two nation-based cultural groups, i.e. where an individual is part of two cultural groups of the same level. One solution would be to assign the individual to one of the groups depending on the context. Due to the inherent stability of Hofstede's system of cultures, his model does not provide a solution for incorporating individuals with values and beliefs merged from two cultures, or for the generation of new cultural forms. Bicultural individuals were therefore excluded from all our studies.

1.1.5 Summary

To sum up, the understanding of culture's impact on the use of Information Systems is gaining increasing importance among scholars and practitioners, yet the current state of cross-cultural IS

[15] Language is, to a certain extent, linked to the concepts of culture and nationality, but has also to be considered seperately from these two concepts.

research is largely challenged by methodological objections. This is reflected in recurrent methodological discussions often preceding empirical analyses. In particular, the strong linkage between the research objectives, an adequate definition of culture, and suitable methodological approaches were discussed. Within this framework, we exposed current debates and clarified our own concept of culture as employed in our empirical research. The methodological section at the end of this chapter presents specific perspectives of quantitative research and its application to the health care field. We also specify our definitions of cultures as applied in our analyses there.

Language plays a particular role when information systems aim at an international audience. It is often considered part of cross-*cultural* research, yet we argue that a number of linguistic aspects go beyond cultural investigations. Language therefore receives particular attention in our research. Its characteristics and role in IS research are presented in the next chapter.

1.2 LANGUAGE

1.2.1 Language as a Research Topic in Information Systems

As discussed in the previous section, there are a number of variables that could potentially determine appropriate market borders. Traditionally, national boundaries have been playing an important role. However, the increasing ease of global communication and international commerce, such as through the Internet, weakens the role of national borders. Other barriers emerge instead. Language, as an easily tangible characteristic that is inherent to people, users, and consumers, seems to represent an ideally distinctive attribute for market segmentation where geographical and national borders disappear.

Yet given the large number of languages spoken worldwide, an adaptation of information systems to all languages is financially and technologically infeasible. Linguistic adaptations of information systems are therefore always restricted to a rather limited number of languages. With regard to the users, foreign language competencies steadily increase as a result of growing internationalization. Consequently, the question is raised as to which languages should be considered for adaptations. To which extent does language affect the use of information systems? Since these linguistic adaptations may represent important financial ramifications, insight into the role of language in IS research is mandatory.

The following paragraphs of this thesis as well as the respective sections of the subsequent chapters investigate the extent to which, and through which mechanisms, language affects how information services are perceived, how it influences website access, and to which extent language has an impact on service and website use and preferences of information presentation.

1.2.1.1 Characteristics of Language

A study of the literature shows that language-related aspects receive far less attention in IS research than cross-cultural issues. Moreover, language is mostly conceived as one among other traits of a cultural group (Hofstede, 1991, p.12; Hölscher, et al., 2005; Reeder, et al., 2004). We argue however, that the role of language goes far beyond its function as an overt cultural attribute. Culture and language are definitely intertwined, but as illustrated in the following paragraphs, several aspects of language are universal or do not coincide with cultural boundaries.

What is language? Whereas the concept of culture is discussed controversially in IS literature, disagreement among scholars regarding the term of *language* is rare.

For this purpose, we refer to one of the early definitions in linguistics, formulated by Sapir (1921, p.8) in order to elaborate important characteristics of language as a medium of communication[16].

"Language is a purely human and non-instinctive method of communicating ideas, emotions and desires by means of a system of voluntarily produced symbols." (Sapir, 1921, p.8)

The key ideas of Sapir's definition of language are the following:

- **Language is a method of communication.** Language serves to exchange information, including thoughts, ideas, messages, knowledge, and emotions. The term communication involves furthermore that the exchange of information occurs between a transmitter who encodes the information and a receiver who decodes it, as stated in models of communication theory (Shannon and Weaver, 1949). With regard to the use of Information Systems, the system provider has the role of the transmitter, whereas the user represents the receiver. In cases of communicational incompatibilities (e.g. different languages (meaning 1 and 2) are used) either party needs to adapt to the language of communication partner for successful communication to occur. This could happen either through the translation of a website or with the acquisition of a foreign language. The role of the market in this context (which is usually determined through the purchase power of the totality of a language's speaker) is one of the thesis' research topics.

- **Language is a system of symbols.** The term *symbols* refers to the fact that words do not have a meaning per se. Meaning is assigned to them and has therefore to be acquired by the speakers of the language. The meaning of a word may shift within one language (e.g. the word "*hound*" did originally refer to any kind of dog; its current meaning is narrowed to a specific kind of hunting dog), the same word has more than one meaning within the same language (e.g. the word *fly* can be the term for moving through the air or describe an insect), or the same word may acquire diverging meanings in different languages (e.g. the use of the word *Beamer* means projector in German, but refers to *BMW* automobiles in (North-American) English). In the structuralistic tradition of linguistic science the term "*system"* relates to a

[16] The definition of the term language as used in linguistics has seen important changes over the years, leading to an increasing formalization and precision Chomsky, N. (1957): Syntactic Structures, Mouton, The Hague, Netherlands.. Consequently, this early definition of language has been criticized by later linguists. Lyons Lyons, J. (1981): Language and Linguistics: An Introduction, Cambridge University, Cambridge, England. puts the question of whether language is purely human and non-instinctive in doubt. He furthermore argues that "voluntarily produced symbols" applies also to body language and gestures which should be excluded from the proper term of language. Finally, Lyons also criticizes the non-exhaustiveness of "ideas, emotions, and desire".

theoretical model where all of the symbols belonging to the same language constitute one entity, and can only be interpreted relative to each other. The meaning of one specific word is therefore limited by words with similar meanings. Accordingly, the linguist Saussure defined each word's meaning in a negative way: "words in themselves carry no meanings, but meaning is communicated by how words (as signs) *differ* from others in the language (or system)" (Cilliers, 1998). For example, the French word *"fleur"* means flower as well as blossom since in French there is no specific word that would allow distinguishing the concepts. Figure 1 depicts a further, well-known example.

<table>
<tr><th>English</th><th>German</th><th>Danish</th><th>French</th><th>Spanish</th></tr>
<tr><td>Tree</td><td>Baum</td><td rowspan="3">træ</td><td>arbre</td><td>árbol</td></tr>
<tr><td rowspan="2">wood</td><td rowspan="2">Holz</td><td rowspan="3">bois</td><td>Leña</td></tr>
<tr><td>Madera</td></tr>
<tr><td>woods</td><td rowspan="2">Wald</td><td rowspan="2">skov</td><td>Bosque</td></tr>
<tr><td>forest</td><td>forêt</td><td>selva</td></tr>
</table>

Originally presented by the Danish structural linguist Louis Hjelmslev, (1943). An extended version from Buckley is presented here (2001).

Figure 1. Structural Relativity in Word Meanings

Differences in languages' structures may become important when the question of how to categorize offered information is raised (see also next paragraph). We address this question in chapter 1.

- **Language is non-instinctive**. Despite the controversy over the instinctive character of language (e.g., Lyons, 1981, p.3), the term "non-instinctive" emphasizes the arbitrary relationship between the symbol and its meaning once more[17]. As a result, the meanings of symbols have to be learned and shared between those who exchange information in order to

[17] Saussure introduced the terms „le signifiant" – signifier and „le signifié" – signified.

ensure successful communication. This aspect of language provides a first explanation as to why the acquisition of a language represents a certain cognitive effort. This is particularly true when a non-native language needs to be acquired. The impact of the cognitive effort is investigated in various contexts throughout this thesis.

It should be noted that for the purpose of this thesis, the term *language* is – despite Lyon's critics (see footnote) - restricted to its verbal use.

In our daily communication the term "*language"* is used rather intuitively in many situations. However, depending on its context it can take very different meanings. Two of them are of particular importance for IS research and will be examined in the following sections. It is used in sentences like: *"The author's language is poetic."* or *"He can speak only one language."*. In the first example the term "*language"* refers to a general means of communication that allows human beings to exchange information. According to biologists, language is in fact what distinguishes humankind most from animals. The meaning of "*language"* can be even further specified in this example: *"language"* classifies a way of using language that is specific to that person. (We will refer to this meaning of language as meaning 1.) Individuals' idiosyncratic use of language is in Saussure's terminology called *"parole"*. The concept of "*parole"* is helpful to explain differences between native speakers of the same language – whatever the nature of the difference is (e.g., proficiency level, domain knowledge, social environment, etc.).

In the second example the term *"language"* refers to a specific means of communication that is used within a specific speakers group, opposing it to other languages that are used in other speakers groups (We will refer to it as meaning 2).

Within this thesis the term *"language"* adopts both meanings, depending on the respective research goal and the context in which it is used. *"Language"* as a means of communication (meaning 1) is predominantly used to explain cognitive challenges in the use of IS. It focuses therefore on the users' linguistic abilities with regard to a specific language or a specific topic field. Examples are English-language websites that offer health-related information. Differences in the cognitive effort required for understanding the information can be found between English native speakers and native speakers of other languages as well as between physicians and patients.

For the purposes of our research we furthermore oppose different languages (meaning 2) and their characteristics. Research about this aspect of language refers mostly to sociolinguistic theories, and is particularly helpful when explaining market-related processes. It therefore predominantly reflects

the perspective of the service or product provider, but also examines users' attitudes towards their native and other languages.

A more thorough elaboration on language's characteristics, mechanisms, and functions allows for the assessment of the challenges of language use in information system products targeted at an international audience in more detail.

1.2.1.2 Language and Culture

Language and culture are the main topics of this thesis. In this theoeretical part they are discussed in different sections which implies to a large extent that language and culture are distinct issues. Yet, whenever *language* is used in its second meaning, the question is raised to which degree language is determined by culture. As stated earlier in this article, we argue that a link between language and culture cannot be refuted. However, depending on the aspect of language considered and on the scholar's academic background, (their perception of) the extent of culture's impact on language differs.

In cultural anthropology, language is usually perceived as part of the characteristics of a cultural group. In Hofstede's onion-model (Hofstede, 1991, p.9), language is depicted as the outer layer of culture, implying that language constitutes (a somewhat visible) part of culture. Cultural values and resulting mental models (see section 1.1.2) are seen as the source of linguistic structures.

Linguistic research focuses with regard to the interrelationship between culture and language on the opposite causality chain: linguistic determinism and linguistic relativism[18] seeks to answer the question of how language creates mental models and affects so people's perception of the social and physical world (see 1.2.2.1.1.5).

In view of IS research objectives, we see three key areas where language is considerably intertwined with culture.

- **Communication practices.** Languages differ in how other people are addressed, how criticism is expressed, or how orders are given. Among the important classifications are those of indirect and direct communication styles. (Gudykunst, et al., 1996; Gudykunst and Ting-Toomey, 1988), and between low and high context communication (Hall, 1976). These

[18] Linguistic relativism is a weaker form of linguistic determinism. It sees language as one determinant besides others.

linguistic differences in style often[19] reflect cultural reality (e.g. the extensive system of politeness and honorific markers in Japanese verbs, adjectives, and nouns mirror their high power distant culture (http://web.mit.edu/jpnet/articles/JapaneseLanguage.html last visited: August 29, 2005). Thus, depending on the cultural background, the same sentence may evoke very different reactions ranging from over to under-interpretation. This is particularly true in situations where non-native languages are used. These aspects of language play an essential role in cross-linguistic and cross-cultural virtual communities where information is exchanged directly between individuals of different linguistic and cultural backgrounds.

- **Lexical structure of the language.** Cultural background (in its broader sense) in terms of physical and social environment also affects a language's lexicon. The impact of culture regards culture-specific words as well, as a lexical elaboration in some vocabulary areas. *Culture-specific words* are words that exist only in a specific language because the item they describe does not exist in other cultures (e.g. "*Shortbread*"). *Lexical elaboration* describes phenomena where a language has several words for designating different forms of an item that has usually only one equivalent word in other languages. Lexical elaboration occurs in the vocabulary areas that are important to that culture (e.g. different Sami words for reindeer, words for rice in many Asian languages - Dirven and Verspoor, 1998, p.145, see also section 1.2.2.1.1.5). The structure of a language's lexicon is assumed to play an important role in the users' preferences of information categorization as well as with respect to their information seeking behaviour (see section 2.3.4).

- **Attitudes towards other languages**. Language is an identity-constitutive factor. Recognition of other individuals as in-group or out-group individuals is based on language traits as well as on (other) cultural traits (Gudykunst, et al., 1989, p. 145). According to socio-linguistic theories, attitudes towards other languages is less an attitude towards the language itself and much more an attitude towards the speakers of those languages, and hence towards their culture (Gudykunst, et al., 1989, p. 147). However, the group assignment occurs through

[19] Correlations between language use and culture are not always trivial and sometimes misleading. Kashima and Kashima Kashima, E. and Kashima, Y. (1998): The Case of Cultural Dimensions and Personal Pronoun Use, Journal of Cross-Cultural Psychology 29, pp. 461-486., for instance, found a correlation between a culture's Individualism score and the use of the word "I". Nevertheless, we argue that the use of the word "I" depends on a number of aspects, in particular efficiency aspects that are inherent to a language's structure. For example, the use of the word "I"/the use of pronouns is mandatory in language were the conjugated verb forms do not allow a phonetic distinction (e.g. French: je mange, tu manges, il mange, etc.) whereas in other languages the verb ending inherently indicates the necessary information making pronouns dispensable (e.g. Italian: mangio, mangi, mangia) (Martinet, André (1962): A Functional View of Language, Oxford University Press, London.. The correlation between culture and language should therefore be used with caution here.

language. In a similar manner, speakers' attitudes towards their own language (especially in smaller speech communities) is largely determined by their identification with their own culture, of which language is a part. In history, language played an important role in the building of a nation, including the promotion of one particular language the and suppression of other languages, language varieties, or even alphabets. (e.g. France, Spain under Franco, The Russian Federation under Stalin - see also Krouglov, 2001; Wright, 2000). Users' attitudes towards their own and other languages is supposed to affect the way how non-native language offers on the Internet are perceived, accessed, and used. In chapter 4 we examine users' attitudes towards a non-native language offering.

1.2.2 The Role of Language for Information System Products

After having discussed the key characteristics of language, the question is raised which role language plays with respect to the use and acceptance of Information Systems. Referring to the mentioned key characteristic we look into single, more specific characteristics of language that are crucial to the questions investigated within our empirical work. The role of language is discussed from two main perspectives: the user perspective and the provider perspective. The distinction is based on whether or not the role of language depends on the interaction of individuals/market players. The user-perspective subsumes aspects of language that can be analysed with respect to one single individual and do not depend on other individuals. For example, reading difficulties depend to a large degree on the language proficiency level of an individual, and change hence of this individual's proficiency level changes, but are not (considerably) affected by other people's changes, behaviours, reactions, etc. The provider-perspective regards mainly market aspects and considers the effect of language as a function of the Internet language offer, i.e. as a function of other provider's language strategies. For example the perception of the use of a particular language on website is examined taking the language offers of other websites into account. As such the provider-perspective considers cost and benefits aspects, relative to the provider's goals.

Language is an important means of communication between human beings. As such it is a necessary precondition for exchanging information and consequently for IS deployment, where information exchange constitutes an essential purpose. The role of language in IS covers a wide field: Language can be used in manuals describing a product's functionality. It can be necessary in order to operate a system (e.g. automated instructions). If the product consists of offering information on the Internet (e.g. online newspaper, health education) language represents an inherent characteristic to the product. It should be noted that on the Internet in particular, the importance of language is augmented since the possibilities for presenting a product are limited

compared to the off-line world: certain categories of information gathering such as touching or personal conversations are not available or are very restricted over this medium. The importance of (written) verbal language is therefore higher on the Internet than in physical stores.

As a consequence of the arbitrary character of language (see above), a shared reference system between information sender (e.g. website designer) and information receiver (e.g. website user) is crucial for successful information exchange. As a result, information exchange is most efficient between members of the same speech community, i.e. between individuals who speak the same native language and have the same cultural background[20].

Websites can be accessed from any place in this world that is connected to the Internet. Due to increasing globalization and internationalization, information systems are used worldwide. The Internet, as an example of information systems in its broader sense, leads to a global information exchange between people of very different linguistic backgrounds. Therefore, in order to allow global communication, website providers either have to adapt their products and services to the users' linguistic preferences, or the users have to be willing (and capable) to view product and service information in the limited number of languages offered by the website providers.

Discussions in the following paragraphs will present arguments why resorting to a language other than the native tongue represents an inconvenience for both providers and users that is only justified if traded off against additional benefits. From the user-perspective these benefits could be, for instance: Higher quality of the provided product or service, better functionality, lower prices, or easier access to information. From the provider's perspective, the major benefit in offering additional languages would consist of reaching a higher number of people, for commercial purposes as well as for societal goals. Nevertheless, a clear separation between the two perspectives is hardly achievable and not intended. Indeed, as can be inferred from our argumentation, both points of view mutually determine each other.

In the two following paragraphs we present different approaches to elaborate the role of language for both the user group and the service-provider group. In view of our research, we predominantly

---

[20] The term "speech community" has been defined by linguists in many different ways. Definitions share "the concern for a definition of speech community as a real group of people who share something about the way in which they use language" Duranti, Alessandro (1997): Linguistic Anthropology, Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge.. In recent publications the term is also for Internet related investigations (e.g., Paolillo, J. (1999): The Virtual Speech Community: Social Network and Language Variation on IRC, Journal of Computer-Mediated Communication 4 [4], p. [online] Available at: http://jcmc.indiana.edu/vol4/issue4/paolillo.html. URL: http://jcmc.indiana.edu/vol4/issue4/paolillo.html.

compare native speakers and non-native speakers. In this thesis the *native speakers* group encompasses all users whose language is offered on the website. The group of *non-native speakers* represents all of the users who have to process information in a language other than their mother tongue. We refer to native speakers also as *L1 users,* and to non-native speakers as *L2 users* [21].

1.2.2.1 The Role of Language from the User Perspective: Cognitive-Behavioural and Attitudinal Approaches

1.2.2.1.1 The Cognitive-behavioural Approach

The term *cognition* is used in various areas in several loosely related ways. Cognition is defined by Preece et al. as follows:

"Cognition refers to the processes by which we become acquainted with things, or, in other words, how we gain knowledge. These include understanding, remembering, reasoning, attending, being aware, acquiring skills and creating new ideas." (Preece, et al., 1994, p. 62)

A cognitive approach to language aims to explain linguistic phenomena with the current understanding of the human mind. Throughout this thesis the term is used to oppose language-related differences between individuals that are due to divergences in language skills to differences that are due to attitudes towards languages. Cognitive and attitudinal approaches to language are somewhat intertwined, but emphasize different parts of how language affects individuals. Although cognition is by some authors considered to be an essential element of attitude (e.g. Rosenberg and Hovland, 1960), other authors point out with emphasis that *"affect"* is the overriding dimension of attitude (e.g. Shaw and Wright, 1967). The latter focus is mostly adopted in language attitude research, in particular in the field of second language acquisition (see also the discussion in Coady, 2001; e.g. Gardner, 1985). Cognitive approaches to language and attitudinal approaches to language are, despite its links, often opposed in linguistic research (e.g. Geerarts, 2003). They are also the subject of different branches of the linguistic field: whereas the cognitive approach is part of cognitive linguistics, attitudinal approaches are the subject of sociolinguistics.

[21] L1= first language, L2 = second language where second language designates all languages that were acquired after the native language(s) without reaching the same proficiency level. Neurolinguistic research further details the distinction between L1 and L2 speakers according to the involved areas of the brain (e.g. Kim, K.H.S.; Relking, N.R.; Lee, K.M. and Hirsch, J. (1997): Disctinct Cortical Areas Associated with Native and Second Languages, Nature 338, pp. 171-174.. For the purpose of our research we deem the first distinction to be sufficient.

It is beyond the purpose of this work to go into the details of cognitive-linguistic processes in the human brain. However, this paragraph gives a first look into the basics of language processing. More detailed insight into the topic can be found in (Hayes, 1970; Tanenhaus and Trueswell, 1995).

Within this thesis, we limit our descriptions of cognitive language issues to the aspects that are central to our research. These aspects are:

(1) Differences in cognitive burden between native and non-native speakers.

(2) Differences in cognitive burden between productive and receptive language use for L2 users.

(3) Differences regarding the time required for information processing.

(4) The correctness of information processing between native speakers and non-native speakers

(5) Differences regarding the semantic network and resulting linguistic differentiation skills.

The cognitive approach towards language is employed predominantly to explain patterns of website access, information seeking behaviour and – as a complementary point of view – patterns structuring information for a website.

Cognitive processes are difficult to measure directly. This applies in particular to the measurement of cognitive burden. The cognitive approach is therefore intertwined with the analysis of behaviour. In fact, within our research we investigated observable behaviour and inferred our insight into the impact of culture and language on cognitive processes from the results from the observed behavioural patterns. This second step is based on the assumption that behaviour is a reaction to cognitive processes: individuals prefer actions that require less cognitive effort (see also discussion in section 2.2.2). A theoretical foundation of the link between cognitive burden and behavioural reactions is provided by the Information Foraging Theory (Pirolli and Card, 1995; Pirolli and Card, 1999) and the Technology Acceptance Model (Davis, 1993).

For the purpose of our research the cognitive approach is furthermore linked to the Information Foraging Theory (Pirolli and Card, 1995; Pirolli and Card, 1999) and the Technology Acceptance Model (Davis, 1993). The Information Foraging Theory identifies link-following behaviour as a resource allocation problem. According to the model, users continue to follow links as long as the information gained from following the link is not exceeded by the costs of accessing it. Cognitive resources required for language processing during information gathering can be regarded as

determinants of access costs and hence link following behaviour. The concept of linguistically determined information foraging is applied within this thesis to the use of search options (sections 2.2. and 2.3) as well as to link following behaviour (section 4.2). With regard to the use of search options, all investigations are also grounded on the assumption that each individual aims at low cognitive effort. Yet the cognitive effort involved with using a certain search option represents a trade-off with inherent benefits of the search option, in terms of time or information quality. For instance, a lower cognitive effort of search option A (e.g. picture supported search) does not automatically mean that every user prefers A over B (e.g. search engine), if search option B offers better features than A, such as faster access to the information sought.

Davis' Technology Acceptance Model (Davis, 1993) (based on Ajzen and Fishbein's Theory of Reasoned Action (Ajzen and Fishbein, 1977; Ajzen and Fishbein, 1980) links users' perceived ease of use with the perceived usefulness, attitude towards using the information system, and its actual use. Language-related cognitive effort affects the perceived ease of use. Therefore, according to this model the costs of language processing affect attitudinal and behavioural outcomes of information system use.

An adaptation of Davis' TAM to a cross-cultural background was introduced by Evers and Day (1997). Their model incorporates *"culturally specific design preferences"* as external stimuli.

Differences in Cognitive Burden between L1 Users and L2 Users: the Revised-Hierarchy-Model

The Revised-Hierarchy-Model by Dufour and Kroll (1995) is a representation of language mechanisms in bilinguals that is often used in cross-linguistic market research (e.g., Luna, et al., 2002; Luna, et al., 2003). Referring to the symbolic character of language (see 1.2.1.1), it describes the lexico-semantic organization of bilingual individuals. According to the model, the lexical representation of a word (*"the word"*, signifier) as a part of language processing has to be linked to its conceptual representation (*"the meaning"*, signified). Whereas lexical representations are stored for each language separately, the conceptual representation is shared across languages. The model predicts furthermore that L2 information processing relies on the individual's first language to gain access to the conceptual representation as a result of the second language acquisition process. The strength of the link between the L2 lexical representation and the conceptual representation increases with improving L2 proficiency. However, residual asymmetry remains between L1 and L2 even after the individual has become fluent in both languages. Consequently, L2 information processing requires more cognitive effort for the user than L1 information processing. In the same manner, L2 information processing at a lower proficiency level requires more cognitive effort than

L2 information processing at a higher proficiency level. Hence, costs of L2 information is higher than of L1 information processing

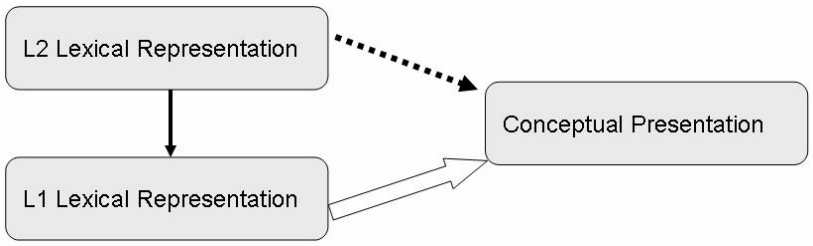


Figure 2. Links between Lexical and Conceptual Presentation in an L1 and L2 Situation (after Dufour and Kroll, 1995)

Differences between Productive and Receptive Language Use for L2 Users

There is further evidence that L2 learners tend to proceed through similar sequences of stages as they gain proficiency with their second language (e.g., Sandra and Rice, 1995) This is particularly true for the acquisition of productive (also: active) and receptive (also: passive) language skills. Listening and reading are referred to as receptive language since information transmitted in the form of verbal language needs to be *decoded*. In contrast, productive language use – which embraces the domains of speaking and writing – requires *encoding* information into the form of verbal language. "Acquiring productive knowledge of a word is a more complex task than acquiring a receptive or passive knowledge of it." (De la Fuente, 2002) Therefore, receptive language skills are usually developed prior to and to a higher level than productive language skills (e.g., Spolsky, 1989). Further evidence for the higher cognitive effort of productive language acquisition and use is provided by the fact that L2 vocabulary development usually progresses from receptive to productive knowledge of a word (Laufer, 1998; Meara, 1996; Nation, 1990). "In most models of L2 vocabulary development, it is assumed that the act of learning a word usually progresses from receptive to productive knowledge." (De la Fuente, 2002) The difference in cognitive effort between receptive and product language is therefore more salient for non-native speakers than for native speakers. Thus, as a result of cognitive efficiency in L2 learning, the retrieval of active vocabulary elicits higher cognitive effort and therefore higher costs than passive vocabulary, especially at lower proficiency levels. Neurolinguistic research has shown that productive and receptive language use also activates different brain regions, namely the Wernicke area and the Broca area (Heilmann and Scholes, 1976).

Differences in Time Requirements between L1and L2 Users

With her study, Hahne (2001) provided evidence that L2 information processing does not only increase the cognitive burden, but it also requires more time than L1 information processing (see also Steffenson, et al., 1979) Differences in time-consumption are mainly due to differences in the degree of automaticity of language processing and consequently in the required resources (further details in Hahne, 2001).

It should be noted however that with regard to time, the relevance of L1 and L2 divergences is also determined by the user's cultural background. The importance of time savings differs between cultural groups, depending on how much time is valued (see differences between short-term and long-term oriented cultures, section 2.3.).

Differences in the Correctness of Information Processing between L1and L2 Users

L1 and L2 users differ furthermore in the fact that L2 users are more likely to misinterpret texts and consequently draw inaccurate information from it (Steffenson, et al., 1979). The relevance of these risk-related aspects depends however on the context again: the characteristics of the product (e.g. price) and its implication for the consumer (e.g. application of health information) are important determinants.

Differences with Regard to the Semantic Network and Linguistic Differentiation Skills

Despite long structuralistic traditions in linguistics, the term semantic network was only first presented in 1969 by Collins and Quillian (1969), as a way in which human semantic memory was organized. According to Anderson and Bower (1973, p. 9) however, the idea of associatively linked concepts is older, and can be traced back to Aristotle. Semantic networks are used in various research areas, such as computer science, psychology, or linguistics. They are usually defined as knowledge representation schemes involving nodes and links (arcs and arrows) between nodes. The nodes represent objects or concepts and the links represent relationships between nodes.

If applied to language, the structure of a semantic network is reflected in the organization of meaning. According to linguistic theories, the organization of meaning can differ between speakers of the same language as well as different languages.

With regard to speakers of the same language, individuals differ mainly in the number of nodes (concepts) that constitute a particular area of the semantic network. The number of nodes determines a speaker's ability to use words with sufficient discriminatory power. Their number and hence differentiation in the semantic network increases with language proficiency (Schur, 2002).

The underlying reason for this development is that cognitive efficiency in language use and acquisition does not allow for redundancy in the beginners' vocabulary (Klein and Perdue, 1992). The level of language proficiency therefore has an impact on the user's ability to use words with sufficient discriminatory power. This ability is important when one has to formulate precise queries, for instance when using a search engine. As a result of the semantic network's structure, users at lower language proficiency levels are likely to enter generic search terms in situations where a precise query formulation would be more appropriate, leading to unsatisfactory outcomes. Research results regarding the impact of differences in language skills between L1 and L2 speakers and between L2 speakers of different proficiency levels on the use of search engines are presented in section 2.3.4.

Differences in the semantic network cannot only be found between speakers of different proficiency levels but also within native speakers of different languages. At the level of the second meaning of language (*langue)*, Sapir and Whorf (Sapir, 1958; Whorf, 1956) infer differences in speakers' thinking and categorization patterns from differences in the organization of languages. The authors suggest that thinking and categorization schemes differ between speakers of different languages due to the linguistic properties of their native languages. According to them and as shown by several empirical studies, linguistically determined categorization schemes lead to different abilities to distinguish or memorize things (e.g., Hardin and Banaji, 1993).

Luria and Vygotsky (1992) give the example of Sami vocabulary with special words

> *"… for reindeer aged 1, 2, 3, 4, 5, 6 and 7 years, twenty words for ice, eleven for the cold; forty-one for snow in its various forms, and twenty-six verbs for freezing and thawing. It is for this reason that they oppose the attempt to make them change from their own language to Norwegian which they find too poor in this regard"*(Luria and Vygotsky, 1992, p.63)

Sapir and Whorf's concepts of the organization of language contain elements that – even though expressed differently – equal a semantic network. The authors explain how the meaning of a word depends on neighbouring terms: the higher the number of words (nodes) in one semantic field, the higher the precision/limitation of one single term[22].

[22] Different levels of distinction regard the lexical field aside from languages' grammatical structure. Sapir described languages lacking categories such as countable and uncountable, noun and verb, tense but endowed with distinctions such as whether an event took place to the North, South, East or West, whether the speaker knew of it from personal observation, from deduction, or from heresay Sapir, E. (1958): Culture, Language, and Personality, Mandelbaum, D.G., Ed, University of California Press, Berkely, CA..

To the second group of languages belong all those languages in which the word can be adequately analysed into a mechanical sum of elements. Each of these elements has its more or less clearly established meaning which is regularly used in all other words into which the associated notion enters.

However, the Sapir-Whorf Theory has a long history of enciting controversial discussions among linguists (e.g., Hunt and Agnoli, 1991). Some of the examples have been proven wrong or not confirmable (e.g. the number of words for snow in the Inuit languages Yupik and Inupiaq[23]). Another difficulty of the theory lies in the question of whether certain abilities to distinguish and remember things are a direct effect of language or whether these abilities simply depend on a person's physical or social environment. For example, a person who regularly has to deal with wine is - independent of the native language - able to distinguish and describe (by the mean of words) a much higher number of tastes and characteristics than a person who deals with wine less frequently. The role of the person's environment here again shows the close link to culture as a determining variable of an individual's perception of the world, and eventually of language (see above). The research field of prototype semantics investigates the role of human cognition for language production and found a number of fundamental similarities across all human beings, independently of their native language. For instance, results from experiments by Rosch (1975) suggest that human beings have very similar underlying perceptive-cognitive categories. A well-known example is the fact the distinction and memorizing of colours is similar between speakers of languages with highly distinctive colour-words (e.g English) and those with more generalizing colour-words (e.g. Dani[24]) (Heider, 1971). Theories of prototype semantics also suggest that so-called *"basic level terms"* are the first and most natural form of categorization across all languages (Lakoff, 1987). *"Basic level terms"* (e.g. table) - in contrast to *"subordinate"* (e.g. kitchen table) and *"superordinate levels"* (e.g. furniture) (Rosch, 1975) allow the provision of sufficiently specific information as well as the inclusion of a high number of subordinate terms (Blank, 2001). They are furthermore usually characterized by simple grammatical/morphological forms (Taylor, 1995).

---

[23] Due to its agglutinative language character, the number of different words depends on the analyst's point of view. *"Agglutinativative"* describes a type of language structure where words are formed by combining/aggregating fixed morphemes. Each of these morphemes has its own meaning/grammatical function. Sapir, E. (1933): Language, Encyclopaedia of the Social Sciences, New York.. (for instance Turkish: *Ev|iniz|de|yiz*: house|your|at|we → "We are at your house.")

[24] Dani is the language of a tribe in Papua New Guinea. Their language only distinguishes between two colours: "mola" for white and all warm colours (red, orange, yellow, pink, etc.) and "mili" for black and all white colours (blue, green) Blank, A. (2001): Einführung in die lexikalische Semantik, Gunter Narr, Tübingen..

Nevertheless, commonalities in categorization do not cover all areas of human cognition and language production. Independently from the fact of whether diverging abilities in distinguishing and describing are a direct effect of a person's native language or an indirect effect through that person's environment, differences in categorization preferences between members of different speech communities can be expected. These differences may become crucial if information is sought on a website. To provide a website structure that allows the user easy and fast access to the information sought, insight into the users preferences of categorization of terms is necessary, in particular if a website designer and website user have different linguistic and cultural backgrounds. Respective research results are presented in section 2.3.5.

Differences with Regard to the Domain Knowledge

It was stated in the previous paragraph that the differentiation of the semantic network is determined by the speaker's proficiency level. In addition, a speaker's proficiency level can considerably differ with regard to certain vocabulary areas. Domain specific knowledge leads to a much higher number of nodes in a particular vocabulary field. Major differences even occur between native speakers of the same language: experts usually have more words at their disposal in their respective vocabulary field than laypersons (see wine example). With regard to L2 acquisition, studies have shown how domain knowledge enhances linguistic performances within specific topic fields (e.g. Goldman and Duran, 1988; Steffenson, et al., 1979; Symons and Pressley, 1993). Furthermore, according to Carell & Wise (1998), the impact of domain knowledge on language skills is more important at lower than at higher proficiency levels.

Consequently, if a speaker's proficiency level is considered to be an important determinant, the speaker's domain knowledge should also be incorporated into analyses as a potential mediating factor. With regard to health related Internet offers, domain knowledge plays an important role if the product or service addresses patients as well as physicians.

1.2.2.1.2 The Attitudinal Approach

Language additionally represents a social value beyond the cognitive effort required of language processing;. The social value of a specific language is based on attitudes towards it, on the part of its native speakers as well as from others. Crystal (1992) defines *language attitudes* as "the feelings people have about their own language or the languages of others".

For its native speakers, a particular language is a constitutive part of a person's identity and cultural environment. As a consequence of its identity-forming role, language presents a distinctive group

characteristic, distinguishing in-groups from out-groups (see below) and carrying group-associated values. Language is, according to Giles & Johnson (1981), one of the most important factors used in social categorization. According to Gudykunst et al. (1989), identifying an individual through language cues as a member of a specific group activates knowledge of the positive and negative values attached to that group and its culture (Gudykunst, et al., 1989). Research also indicates that identification as an in-group member leads to a more favourable perception (Hogg, et al., 1984). This insight into the social role of language is helpful for understanding and predicting user behaviour and attitudes. Favouring a native language from some of the users lies in its characteristic to evoke in-group based positive values (Tajfel, 1978) that are subsequently ascribed to services and products and their providers. Consequently, identification with the native language, beside cognitive processes, is a further cause why L1 information presentation may reduce the users' perceived risk and increase trust towards the product or service. Section 3.3.1 addresses this question.

Under certain circumstances the use of non-native languages also induces associations of positive value, and enhances so the perception of a product or service. In marketing studies, language-associated values contribute to the so-called *country-of-origin-effect*. Here evaluations of products and services are based on values associated with the respective country and/or language (e.g. reliability, trustworthiness, level of technological advancement, etc…). A list of attributes associated with the use of certain languages in international advertising can be found in (Dmoch, 1997). In line with the Elaboration-Likelihood-Model[25] (Petty and Cacioppo, 1986), research indicates that product and service evaluation is only more likely to be directed at country-of-origins-effects if the consumer's motivation is low (Gürhan-Canli and Maheswaran, 2000). Although literature lacks analogous studies focussed solely on language aspects, it can be assumed that language as part of a country's characteristics induces comparable evaluation mechanisms.

Reasons for the development of values that are associated with other languages or the mother tongue are of a very different nature. They range from historical events to economic conditions (Fishman, 1999). A myriad of socio-linguistic studies researched these factors, for instance with the objective of providing powerful incentives for the preservation of lesser-used languages (Fishman, 2001). It is beyond the purpose of this thesis to exhaustively review these determinants. We instead focus on one particular reason that is deemed to be the most important determinant of users'

[25] The Elaboration-Likelihood-Model claims that there are two ways of making decisions. The so-called central route is taken in the case of high motivation. In other cases the peripheral route is taken. Decision-making here is not based on persuasive arguments but is swayed instead by secondary characteristics.

attitudes towards native and non-native language use in service and information offerings on the Internet.

Research results suggest that the size of a speech community is correlated to the acceptance and command of other languages to a certain extent (Dalmazzone, 1999): smaller speech communities are usually more willing and more able to understand and speak other languages. It can be expected that language acceptance (considered a consequence of language skills as well as independently from it) is mirrored in individuals' attitudes towards native or non-native language offers on the Internet.

In line with these assumptions, members of larger speech communities with an extensive supply of products and information in their native language outside the Internet should be less willing to use non-native language services and products on the Internet. An extensive supply of websites in the native language is usually the reason for sparse contact with other languages on the part of the consumer, leading to a cautious attitude and lack of foreign language skills (e.g., Lambert, 1974). With regard to smaller speech communities, the opposite might be expected. Attitudes towards other languages may differ considerably between and within smaller speech communities. In fact, sociolinguistic studies found evidence among smaller speech communities supporting a higher acceptance of non-native language use and a more positive attitude towards it (e.g., Giles and Byrne, 1982; Ljung, 1986) as well as the opposite case (Hall and Gudykunst, 1986).

Nonetheless, these contradictory findings can be explained with the dilemma to which smaller speech communities are exposed. On the one hand, smaller speech communities suffer due to their size, from market constraints that force them to promote trade and communicate with other speech communities, renouncing the use of their own language. On the other hand, native speakers of smaller language groups might exhibit a stronger identification with their native tongue, resulting in a stronger awareness of language issues and a stronger willingness to preserve their language as part of their cultural heritage. As a consequence, resistance to accept the use of other languages can be more salient in smaller speech communities than in larger speech communities.

> *"… the speaking of a minority language that reduces the scope for economic exchange may persist because of strong social customs in favour of speaking the language." (Akerlof, 1984)*

The Internet as a global medium and consequently multilingual setting constitutes a rich basis of research about language attitudes. Considering Internet users' foreign language competencies and the dominance of English websites, the multilingualism of the Internet can be reduced in many cases to a bi-polar language choice: the choice between a website offer in the user's mother tongue

and one in English. Results regarding language-related link-setting and link-following behaviour presented in chapter 4 confirm the dominance of English.

Literature however largely lacks results regarding attitudes towards languages offered on the Internet. In Kralisch (2003), the relationship between the widespread use of English as a native or second language and the adoption of the Internet in its early years was suggested. Despite the parallel impact of other variables such as telecommunication costs, technological advancement, or communication alternatives, these findings provide first indications of the role of language for information consumption on the Internet.

With the research presented in chapter 4 we provided a further contribution to fill this gap. We studied the extent to which Internet users are accustomed to find information about specific topics in their native languages. It was furthermore examined to which degree this attitudinal value of a language is traded against other benefits such as better information service.

1.2.2.2 The Role of Language from the Service-provider Perspective: Market-driven Approaches and Societal Goals

1.2.2.2.1 The Market-driven Approach

Translation for products and services is restricted to a few languages compared to the number of languages spoken in the world. This paragraph examines potential reasons behind language selection for information presentation on a website, on the part of the product and service providers.

Costs-Benefit-Trade-Offs. A market-driven approach to language in its simplest form assumes that the amount of language-related effort by either the provider or the consumer is regulated by market forces. Analyses of underlying economic mechanisms in the use of language have so far provided insight with regard to job market, language policies, or language revitalization (Arcand, 1996; Breton, 1998; Chiswick and Miller, 1995; Dorian, 1981; Drinkwater and O'Leary, 1997). Here again literature unfortunately lacks studies that explore the economic principles of language use with regard to the Internet.

As illustrated in the previous paragraphs, offering services or information in the users' native languages represents an additional service to the customer. Market-based language strategies are founded on the fact that inherent and associated characteristics of language differentiate the services offered and information from others: native language use may reduce the users' cognitive effort and increase the appeal of the website, generating trustworthiness and comfort (see above). From the

service-provider perspective, language can therefore be understood as an instrument designated to increase the acceptance and use of any form of information system product or service.

Yet a linguistic adaptation of the website inevitably involves additional costs. Implied costs for translation and adaptation may differ between languages due to diverging availabilities of translators and language software, as well as to differences in involved technical adaptations (e.g. character sets for Chinese, Japanese). Language-related costs also differ between products depending on their communication requirements (e.g. instruction for computers vs. instruction for shoes - Harris, 1998). However, the fact that outside the Internet language intensive goods are more likely to be localized than goods with low communication requirements indicates that the benefit side of linguistic adaptation (i.e. additional customer/purchase power) is more important than the cost side.

What are the benefits? The most important benefit from offering services or information in an additional language is the increased number of Internet users that can be reached. If language is simply considered a medium of information transmission, the range of a specific language is defined by its overall coverage, determined by the number of native speakers as well as by the number of (potential) Internet users who are L2 speakers of that language. One of the reasons for the widespread use of English on the Internet is indeed not because of the high number of native speakers but the high number of Internet users mastering English as a second language. As an example, Figure 3 provides statistics concerning the language situation in Europe in September 2005. Europe was chosen since reliable numbers for English language proficiency worldwide are not available. Due to its linguistic (and cultural) diversity, Europe can however be considered an illustrative example.

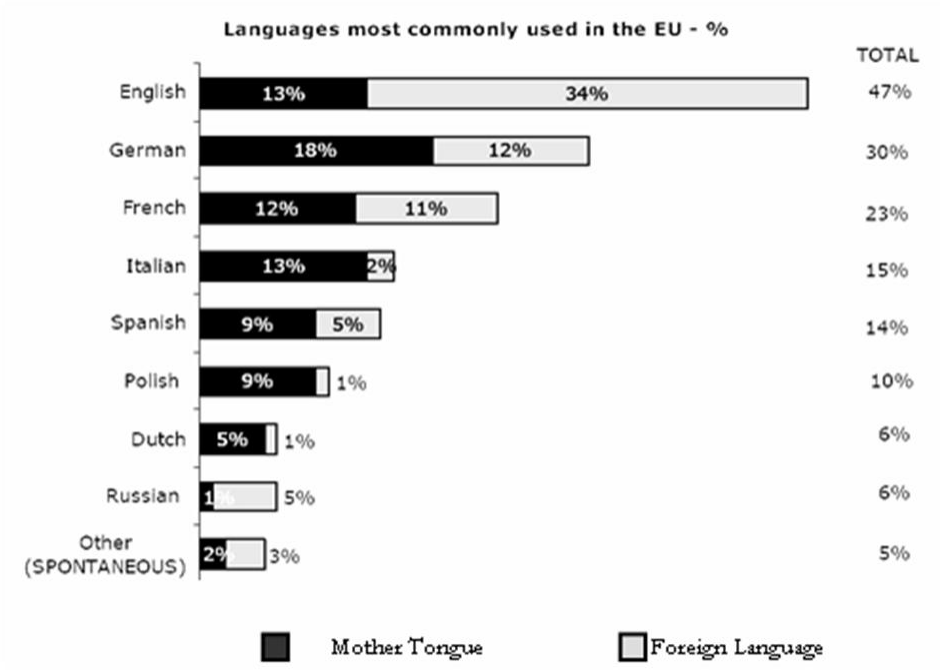


Figure 3. English as Mother Tongue and Foreign Language: Language Use in the European Union
Source: (2005)

As illustrated in the previous sections, a distinction between native language and non-native language becomes important once language is considered beyond its functions of information transmission. Besides the decreased cognitive effort, the social role of language leads to the fact that a native language offer also has a positive effect on the users' attitudinal perceptions of the website in the vast majority of the cases. As a result, it can be expected that users would choose the one that presents information in the users' native languages when choosing between two websites with identical service.

Such an effect seems to be common sense. Moreover, identical services are more a product of laboratory settings than a real world setting. The importance users ascribe to their native languages would be much more revealed if traded off against other features such as price, quality, or choice. Research is definitely required with regard to this field.

Language Markets. The term "language market" refers to the fact that language supply – similar to classic products – can be regulated through market mechanisms of supply and demand. The term is commonly used in studies about the economics of language (e.g. Harris, 1998). The previous paragraph analysed reasons explaining why the size of the speech community has an impact on the users' attitude towards their own and other languages. With regard to market

strategies, the above-listed arguments impose that in a language market with an already existing extensive supply, a native language offering appears to be a precondition for a successful market penetration. In fact, with growing supply, the internet user's power to selectively chose between the service and information offers increases (Bronold, 1999). Hence, independently from foreign language proficiency skills, a non-native language offering strongly risks to be perceived as lack of customer service.

In a language market with a low supply of native language offerings, market pressure to offer a native language website – usually through both competitors (push) as well as customers (pull) – is less salient. Nonetheless, as Grin points out, offering a service in the customer's native language may represent a central marketing differentiation strategy in this case (Grin, 1994, p. 286). Using the native language is a service feature that distinguishes the product or service from competing offers. This strategy is particularly effective in cases of small speech communities with a high attachment to their native language. It is also useful in markets where competition is keen on other, non language-related aspects (Dickson, 2000). It should be furthermore noted that with the growing distribution of the Internet, an increasing number of people lacking foreign language skills participate in this medium. A more extensive list of factors that determine the additional benefit of offering the user's native language is provided by Grin.

However the classification of a language as a lesser used language or smaller speech community is relative. In particular with regard to market-driven approaches, the importance of language also depends to a large extent on the market's boundaries. In fact, the often-cited dominance of English and other widely spread languages such as German, Japanese, or Spanish regards the entity of Internet users as a target market. In diverging market scenarios with a limited number of target groups, countries, or languages the importance and dominance of generally lesser-used languages may be considerably increased[26]. In those cases, a linguistic adaptation to these lesser-used languages may lead to higher benefits since the speakers of these languages represent a larger market share. In a study conducted in the beginning of 2001, results from an analysis of market language strategies of SMEs appear to confirm the impact of market delimitations on the selection of the languages chosen for linguistic adaptations: companies that targeted at a multitude of linguistic user groups usually chose English and the official language of the country where the

[26] It should also be noted that minority languages may increase in importance if the Internet bridged geographically dispersed groups (Warschauer, M.; El Said, G.R. and Zohry, A. (2002): Language Choice Online: Globalization and Identity in Egypt, Journal of Computer-Mediated Communication 7 [4], pp. [online], Available at: http://jcmc.indiana.edu/vol7/issue4/warschauer.html [Retrieved: November 1, 2005]. URL: http://jcmc.indiana.edu/vol7/issue4/warschauer.html).

company was based as languages for information presentation. Companies targeting users of specific target groups much more frequently offered additional languages beside English (Kralisch, 2003).

Nevertheless, despite these considerations it seems that market-driven language choices on websites inevitably lead to an under-representation of smaller speech communities. Even though differences might be less pronounced in a long-term view between large and middle-sized speech communities, small speech communities under a certain threshold are likely not to be served in their native tongue by commercial enterprises. Unfortunately, detailed examinations of such linguistic evolution with regard to the Internet have been only a very recent subject (e.g., Special Issue on Minority Languages, Multimedia and the Web in the New Review of Multimedia and Hypermedia 2005(2), and often disregard commercial aspects.

In our research we investigated the applicability of market and network theories to language use on the Internet. Results are presented in chapter 4.

1.2.2.2.2 Societal Goals

Societal goals in Internet services are characterized through non-profit maximizing purposes. The impact of market constraints is therefore less salient. The most frequent goal of this type of product and service is for educational purposes. Indeed, the Internet is increasingly used as a medium for presenting public information, mainly on behalf of public institutions. For example, thanks to the comparatively low costs implied in Internet publications, public health care education is realized online to an increasing degree.

Similar to market-driven language selection, the main objective of websites with educational purposes is to reach the largest number of Internet users possible. This can include Internet users worldwide or within a specific region. In contrast to rather commercially-oriented offers, information offers of this type also have the responsibility of serving underrepresented language groups. Offering these languages helps speakers of the languages strengthen their position and contributes to the preservation of their language (Warschauer, 1998).

Nevertheless, it is a matter of fact that non-commercial offers are developed and maintained under cost-restraints. Moreover, resources are often insufficient to make use of more expensive media (Warschauer, et al., 2002). Linguistic adaptations are therefore restricted here again to a limited number of languages. This dilemma poses a challenge in particular for public information offers that aim at a large audience. Consequently, the question is raised as to which linguistic factors

determine whether an Internet offer is accepted and used by these underrepresented groups. To which extent does information presentation in a non-native language constrain information comprehension? To which degree does the use of their native languages significantly increase the success of these offers among underrepresented language groups? Thus, contrary to market-driven approaches which focus on what we defined as "the social role of language", language strategies for websites with societal goals emphasize the cognitive consequences of language use.

Nevertheless, a clear distinction between the two website types does not exist. With respect to health information for example, a growing commercialization can be observed (e.g., 2005). Similar developments can be assumed for other areas. Furthermore, the perception of language use for public information offerings is also subject to users' attitudes towards non-native languages, depending on the market situation - which is however usually less competitive for this type of website.

1.2.3 Summary

Two meanings of the term "language" were introduced: *"language"* as a means of communication in general, (meaning 1), and *"language"* as a means of communication specific to one group (e.g., Russian vs. English, meaning 2). Various characteristics of language were described, in particular the (relative) arbitrarity of the relationship between a word (signifier) and the concept (signified). It was stated that some concepts seem to be language-independent whereas others are language-specific and affect an individual's mental models. Language and culture share therefore the characteristic that they affect the development of mental models. Moreover, the culture and language are somewhat intertwined which is reflected in certain lexical or grammatical structures. Language, especially when considered as an overt characteristic of culture, is also an identity-constituitive factor.

In order to investigate the role of language with respect to the use and acceptance of Information Systems we distinguished between a user and a service provider perspective. The user-perspective subsumes a cognitive-behavioural approach and an attitudinal approach. Cognitive and behavioural aspects or linked within one approach due to the fact that cognitive processes are best and easiest measured through analysing behavioural actions. The discussion of cognitive aspects of language revealed an easier, faster and correcter language processing for native speakers than for non-native speakers. We furthermore pointed out that productive and receptive language use and domain knowledge lead to further differences in language processing.The attitudinal aspects of language are closely linked to language as a cultural, group-specific characteristic or/and identity-constitutive

factor. An individual's attitude towards a language was identified as a further determinant of language-related use and acceptance of Information Systems that is relatively independent of cognitive aspects.

Insight into these user-focused aspects of language are also important from the service-provider perspective. Within the provider perspective, cost and benefit aspects of offering a specific language were considered. A provider's benefit of offering a certain language depends largely on the number of potential users, but also on to which degree users (would continue to) accept a non-native language offer. This, in turn, depends on the effects of language that were described within the user perspective (cognitive and attitudinal aspects), but also on other provider's language offers. This is because the availability of language might change the attitude towards a non-native language offer and might affect over a long time also the users' non-native language skills.

A similar reasoning can also be applied to cultural adaptations. Nevertheless, cultural adaptations are less visible and more gradual and therefore less perceivable to the user. The role of the *availability* of culturally adapted websites on attitude formation/acceptance seems therefore to be less important. However, the impact of culturally formed mental models can be expected to considerably affect use and acceptance of Information Systems and thus the benefit of a culturally adapted website for the service-provider.

Chapter 2, 3, and 4 investigate through empirical studies selected questions of how native language offers in contrast to non-native language offers and culturally formed mental models lead to certain behavioural and attitudinal reactions.

In preparation of the empirical part of this thesis, the following section reviews the quantitative approach adopted for our research and presents applied methods, the conceptualization and quantification of culture and language and gives an overview of participants and material used for our studies.

1.3 METHODOLOGY: METHODS AND CONSTRUCTS

1.3.1 The Appropriateness of Quantitative Research for Cross-Cultural and Cross-Linguistic Studies

1.3.1.1 Introduction

In the two previous sections we presented an extended overview of the thesis' two main research subjects. As illustrated, methodological approaches for investigating cultures are fervently discussed. Quantitative cultural studies have especially been more and more strongly contested in

recent years. With respect to language, methodological debates seem to be less relevant. Nevertheless, depending on the study's content, different research methods are applied. Whereas cognition-related analyses are almost exclusively based on quantitative studies, the social role of language is investigated by both quantitative and qualitative methods.

Our research designs are predominantly based on quantitative methods. The reasons behind this choice are varied, and are presented in the two following sections. It is deemed beyond the purpose of this thesis to provide a detailed picture of epistemological and ontological discussions in IS research. Yet, an extended introduction to the subject helps depict reasons and underlying assumptions regarding the dominance of quantitative approaches adopted in our research. In the following paragraph we therefore want to give a simplified picture of what distinguishes qualitative approaches from quantitative approaches. In this context, advantages and drawbacks of the research approach for cross-cultural and cross-linguistic IS research are discussed.

Methodological debates about the appropriateness of qualitative and quantitative approaches in IS research have always been influenced by this application orientation and multidisciplinarity. Its multidisciplinarity and application orientation give rise to a variety of research methods and theoretical foundations borrowed from other disciplines, each of them characterized by their own particularities and research traditions.

The ongoing discussions presented next, and also our own approach, therefore have to be interpreted and evaluated taking this context into account. In particular, the fundamental relationship between research objectives and methodological appropriateness is essential to our argumentation line.

1.3.1.2 Epistemological Assumptions of Quantitative and Qualitative Research

One possible distinction of research methods is between quantitative and qualitative research. However, giving a precise definition of each term is not an easy task. Quantitative and qualitative research are usually defined in a negative way (such as "quantitative research's findings are not arrived at by statistical or other quantitative procedures") or by comparing their research goals and the applied research methods. For example, qualitative research is usually described as less structured than quantitative research. As such it employs research methods such as observation, content analysis, (unstructured) interview, case studies, or focus groups.

This section aims to provide more information about the distinction of these two research methods and the characteristics of each of them. It puts qualitative and quantitative research methods into

the context of epistemological objectives and the research disciplines where they were traditionally applied.

Clarke (2000) distinguishes between three research approaches appropriate for IS research: quantitative research, qualitative research, and engineering research. The author classifies the first two approaches as business disciplines, as opposed to engineering disciplines and computer science (Clarke, 2000). Clarke himself avoids the terms of quantitative and qualitative research and replaces them by conventional scientific and interpretivist research. This lexical choice first of all insinuates the traditional role of quantitative methodological approaches in scientific research, and is at the same time an indicator of the ongoing debate about a distinction between the two approaches.

Despite many years of fervent discussions, debates about qualitative and quantitative methods cannot provide any general insight about the usefulness of one of these methods. Their appropriateness can only be evaluated taking the study's goal into account. Judgements about the suitability of qualitative and quantitative research are based on underlying epistemological assumptions (Trochim, 2000) which simply can differ from researcher to researcher without one being superior to the other.

Quantitative studies emanate from natural sciences and were increasingly adapted to social sciences only in the last few decades (Stangl, 1989). Their epistemological objective consists in obtaining "objective knowledge" and "true facts" about the world (Lienert, 1989). Research is based on a large number of single observations, conducted to confirm theories and hypotheses that were postulated beforehand. Thus, insight in quantitative studies is gained through confirmatory, inductive research processes (Clarke, 2000; Trochim, 2000) which "… commences with observation of specific instances, and seeks to establish generalizations" (Hyde, 2000; Lienert, 1989).

In contrast, assumptions underlying qualitative research reject the existence of objective facts and truth. Epistemologically, the goal is to understand a phenomenon, a research objective requiring analyses that take a large variety of context into account. Its comprehensivity is the reason why qualitative research has traditionally been playing an important role in social sciences (Hoepfl, 1997). Observations are consequently conducted in a rather exploratory way. Deductive conclusions are more often applied in qualitative research than in quantitative research, although also qualitative research most often follows an inductive process (Hyde, 2000). "Deductive reasoning commences with generalizations, and seeks to see if these generalizations apply to specific instances." (Hyde, 2000)

Adherents of the qualitative approach often criticize the narrowness of assumptions in quantitative research. Indeed, quantitative research is limited in nature and can only illustrate a small portion of reality (Trochim, 2000). The necessary quantification of constructs and variables inevitably inheres a reduction of complex processes and coherences. This reduction can however be interpreted as a type of standardization allowing the comparability of research results.

In contrast, research results from qualitative approaches have little *validity* as defined by quantitative research approaches. Due to the specific contextuality of each research situation, replications and comparability can hardly be achieved. Hence, in its simplest form, the question of quantitative or qualitative research is the question of whether simplifying and standardising effects through quantification are aimed at, or whether revealing the processes of a complex phenomena is emphasized.

Besides, it is often argued that each qualitative approach contains quantitative elements and each quantitative approach embodies qualitative parts. One example is that analyses of words - as they often occur in qualitative research - can and are usually coded quantitatively. Conversely, numbers as used in quantitative research are based on qualitative judgement and several underlying assumptions (e.g. Did the respondent understand the meaning of the numbers in the survey?) (Trochim, 2000). Also, in most instances quantitative theory developed from qualitative investigation of untested theories (Hyde, 2000).

Finally, given the topic of this thesis, the extent to which favouring either quantitative or qualitative research approaches is biased by the scientist's cultural background should be questioned. Although speculative at this point, the dominance of quantitative studies in sciences shows indeed a relationship with the dominance of Western culture in science, and its preference for linear rather than comprehensive thinking inherited from the Greeks. "Studying the model designed to explain nature tells more about their creators than about the part of nature being studied". (Missana)

1.3.1.3 Appropriateness of Quantitative and Qualitative Research in Cross-Cultural and Cross-Linguistic Studies

After having situated qualitative and quantitative research in the context of epistemological objectives, one needs to answer the question of how appropriate these methods are for the purpose of cross-linguistic and cross-cultural IS research.

Cross-linguistic and cross-cultural IS research is close to various social science disciplines and psychological studies. It often borrows its theoretical basis and methodological approaches from its

reference disciplines. In particular, cultural studies are often conducted employing qualitative approaches such as ethnographic research (Honold, 2000), repertory grid technique (Agourram and Saucier, 2004), or in-depth interviews, etc. (Gillham, 2005). Indeed, examinations of culture's impact appear to be prone to be conducted with qualitative techniques for a number of reasons. First, as can be inferred from our discussion in section 1.1.3, culture is a particularly complex and context-dependent construct. Additionally, the quantification of cultural data embodies a considerable reduction of culture's idiosyncratic complexity, and also requires a number of assumptions. Finally, the lack of quantitative data alternatives limits the variety of quantitative cultural research.

We argue that cross-cultural studies and cross-linguistic studies in the context of IS research differ considerably in their objectives from classic social studies due to their application orientation and integration into a business and engineering environment. Business and engineering areas have a rather strong tradition of quantitative research, which finds its justification if the environment of decision processes and product development cycles is considered. Cross-linguistic and cross-cultural studies are the basis for investment decisions regarding product or service adaptation in a business context. Investment decisions are not usually made by experts of cross-cultural and cross-linguistic differences. Information regarding cross-linguistic and cross-cultural adaptations therefore need to be acquired by the decision maker. Also, information needs to be understood within a short amount of time. Quantified results help rationalise the process of decision making by presenting a reduced picture of reality in the form of "numbers" and "facts". The question of how useful simplified pictures of reality are for this kind of decision remains, and to which degree reality should be reduced.

Also, the implementation of language and culture related to the product design is more challenged if insight is obtained from qualitative research. Various field reports describe, for instance, the difficult information transfer between ethnologists and programmers (Kirah, 2005). Thus, quantitative approaches in cross-cultural (and cross-linguistic) IS studies are predominantly justified through the studies' integration into investment and product development cycles, requiring matching approaches between the two disciplines.

Finally, it should be noted that the quantitative approach also resulted from the proximity of our studies with psychological research, where inductive research has a long scientific tradition and high level of recognition.

1.3.2 Applied Methods

The dominance of quantitative approaches in our research implies an important epistemological premise of this work: we emphasize testing well-known theories from other disciplines, applied within the context of cross-cultural and cross-linguistic Internet communication. In particular, theories from psychological research (such as the Theory of Reasoned Action (Ajzen and Fishbein, 1977; Ajzen and Fishbein, 1980)) were concerned. In contrast, the exploratory identification of undetected relationships and regularities and the understanding of single phenomena received less attention in our studies.

In order to obtain data for answering the research questions outlined, we gathered data through different methods. These methods allow the collection of data that is suitable (after transformation) for statisctical procedures. Nevertheless, the methods differ in their appropriateness for investigating specific theories. In order to compensate for the limits of one particular method, in the majority of our studies we combined two or more methods. In particular, high external validity was assured through employing logfile analyses, whereas higher internal validity is provided (to a certaint degree) through questionnaires or laboratory studies.[27] The particular characteristics of each data gathering method employed are illustrated in subsequent sections of this chapter.

A large part of our findings are based on logfile analyses. The discussion of this methods' suitability is preceded by brief introductions due to the fact that the application of logfile analyses is rather unique and little known in cross-cultural and cross-linguistic research.

1.3.2.1 Logfile-Analyses

In our research we mostly relied on logfile-analyses. Logfile analysis examines the information logged on a server by users' website visits. We analysed in the relevant studies data concerning one single website. In study 8 we added information from a web crawler in order to obtain additional information that allowed us to put the concerned website into the context of the World Wide Web. The web crawler provided data about how other websites link to our website, the target website. Further details are presented after the discussion of aspects of logfile analysis, in section 1.3.2.1.3.

---

[27] Internal and external validity, in addition to construct validity and statistical validity, are the main forms of validity distinguished by Shadish, Cook, and Campbell Shadish, W.R.; Cook, T.D. and Campbell, D.T. (2002): Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Houghton Mifflin, Boston, MA.. Internal validity regards the question of whether the research design is appropriate to provide evidence for the correlation between the independent and the dependent variables. External validity relates to a possible generalization of the research results to other groups than the tested sample group.

Data from logfile analyses provide insight into the behaviour of the website's users, and allows, to a certain extent, conclusions about the user's preferences and difficulties when navigating on the website. The logfile data contains various information. In our studies we mainly refer to the user's IP address, the requested page, the page that was requested before (referrer page), and the time of access. Additionally, a particularity of the specific logfile we used allowed us to obtain information about the language in which the website was visited (lang=…). Figure 4 gives an example of a logfile as employed in our studies. Each requested page (click) is recorded as one entry (line) in the logfile.

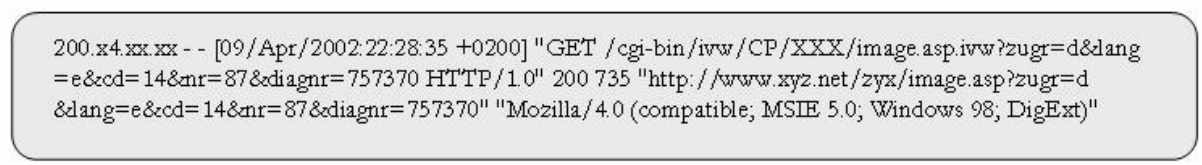


Figure 4. Example of Logfile Data

Logfile analyses constitute a data gathering method that is appropriate for obtaining highly structured data that allows for statistical analyses. We primordially obtained the number of occurrences of certain events (e.g. request of a particular page (set)) within one session or within an aggregated group of individuals (e.g. group of native speakers) from logfile data. Time consumption was a further often-collected piece of information. In order to obtain structural information, i.e. information about the form of navigation sequences (see study 1 – section 2.3.2), preprocessing data analyses and data structuring were required (e.g., application of taxonomies). In order to make logfile data available for statistical procedures, we transferred the data set into a MySQL or Oracle database. Data was then analysed or further processed by means of queries. Final statistical analyses were done with SPSS 11.0 and 12.0.

Finally, it should be noted that data Mining in general, and logfile analysis in particular, can be conducted either in an exploratory or in a confirmatory way. As already suggested by the definition above, exploratory knowledge discovery is a traditional approach in data mining. A classic example is the detection of frequent patterns within the data set. However, confirmatory models based on the knowledge gained from data mining, tested for validity in another sample set, often complement this exploratory approach. Examples of data mining applications are the fraudulent use of services (e.g. for credit card use, insurance companies) or for marketing purposes in business areas, including customer profiling and cross-selling (Berry and Linoff, 1997). Within our studies, logfiles analyses were employed for confirmatory approaches.

1.3.2.1.1 Advantages and Limits of Logfile Analyses

In our evaluation of the suitability of logfile analyses for cross-cultural and cross-linguistic IS research, we refer to two key quality criteria: the control of variables (internal validity) and the generalizability of results obtained (external validity).

One of the advantages of logfile analyses is its accommodation of the large size of the data set. Such a large sample set offers unprecedented opportunities in terms of the amount and variability of data and users studied. Necessary assumptions about the equal distribution of specific variables are more justified if the sample group is large. Hence, logfile analyses usually ensure high external validity and therefore the generalizibility of the study.

Yet, despite the large data set provided by a server's log, logfile analyses do not allow to control for specific variables. Logfile data assesses behavioural variables, but data about the users' attitudes and reasons for their behaviour cannot be obtained. Splitting the user-group into two groups with different conditions (treatments) and gathering data about these two user groups is difficult if the analysed logfile comes from a real website. The internal validity of a logfile analysis is in consequence usually low and can only be assured in combination with other methods.

Nonetheless, external and internal validity cannot be considered independently of each other. First, practical reasons usually require the researcher to focus on one form of validity. Second, an increase of external validity may also increase the internal validity and vice versa. In our case, a large variance of non-cultural and non-linguistic variables within each cultural or linguistic group increases the chances that the cultural and linguistic variables investigated are not confounded with other (e.g., socio-economic) variables, unless systematic relations with the cultural or linguistic variable exist. As such, the correlations found are much more likely to be produced by the impact of culture and/or language.

Ecological validity is one specific aspect of external validity, regarding "the extent to which results can be generalized from the set of environmental conditions (in laboratory studies created by the researcher) to other environmental conditions (settings and conditions)" (Bracht and Glass, 1968). Logfile analyses belong to the group of non-reactive data gathering methods, i.e. analysed test-subjects do not notice the test-situation. Non-reactivity is one way to ensure that the behaviour analysed behaviour does not differ considerably from the user's usual behaviour (Albrecht, 1982; see also Albrecht, 1985; Bungard and Lück, 1974; Webb, et al., 1981).

A further reason why logfile analyses are suitable for assessing genuine behaviour consists of the fact that data from logfiles usually record behaviour in the user's authentic environment. As a result, it can be furthermore assumed that the users analysed are typical representatives of their cultural or linguistic groups, in contrast to recruited international students. "The extent to which the results of a study can be generalized from the specific sample that was studied to a larger group of subjects" -with potentially diverging characteristics is referred to as population validity (Bracht and Glass, 1968), which is consequently large in logfile analyses.

Finally, since logfiles assess data about every single user, the bias of self-selection is considerably reduced and solely regards the exclusion of non-users. Thus, logfile analyses are characterized by high ecological but also high population validity that may not be reproduced in a laboratory setting (see also Reips, 1999).

Nevertheless, a server-centric design that only allows an examination of the website users (vs. non-users) represent some of the limitations of our methodological approaches. Firstly, it does not allow the gathering of important data about people who do not visit the website for some reason. Secondly, it is not clear to which extent insight is limited to the specific features of the investigated website.

Furthermore, unless persistent cookies are used, analyses can only be carried out on the session level. However, while employing persistent cookies would make it easier to identify unique website visitors, it might also bias the results due to the effects of the privacy concerns of many users who disable cookies. The extent to which such a data aggregation biases results depends on the study's objective. In some cases, an analysis on the session level can be justified by the fact that frequent websites visits by the same user reflect a low access barrier or a high usefulness.

After all, it should be also noted that logfiles are an easy means of data collection but require laborious data preparation and data analysis. In addition, data preparation and analyses often rely on heuristics that underlie several assumptions. The most important one is presented in the next paragraph. As is common in other studies, we assumed the size of the sample to outweigh this noise. In a few cases, the impact of applied heuristics on the data set is documented in the following paragraph.

1.3.2.1.2 Application of Logfile Analyses in our Research

Despite the number of positive aspects, the use of logfile analyses for the purpose of cross-linguistic and cross-cultural studies is still rare. Exceptions are, for example, Stander et al. (2004)

and Komlodi (2005). In our research we employed logfile analyses differently from its traditional use. Whereas data mining usually follows an exploratory approach, we predominantly used logfile data in order to find evidence for our postulated hypotheses. This use of logfile analyses as a confirmatory methodological tool is in line with our inductive research approach.

Studies based on logfile-analysis. Logfiles are the main data source in the studies where we investigated users' navigation patterns and search behaviours with regard to linguistic and cultural differences, including differences in domain knowledge (see Chapter 2). Measures used for each analysis are described in their respective sections.

We furthermore used logfiles to determine the extent of users' website use (e.g. numbers of page requests) as an additional measure to questionnaire based studies (see 2.3.4). Finally, logfiles were also one of the data sources for examining link patterns (predominantly based on referrers) and user flow (see Chapter 4).

The logfiles themselves were provided by two health content websites, which are described in subsequent paragraphs (see 1.3.4).

Data Preparation and Cleaning. In order to allow the analysis of user behaviour, two essential steps of data preparation are necessary: session reconstruction (sessionizing) and data cleaning. As mentioned above, one entry line in a server log represents one page request. Entries are ordered by time and not by users. This explains why each entry has to be assigned to one user ex post. Sessions that describe each user's navigation path through the website are inferred from the logfile by using information about IP addresses, time data and data about the referrer page (e.g., Cooley, et al., 1999). Generally, sessionizing requires employing certain heuristics, which by nature cannot assure a perfectly realistic reconstruction of the sessions. Nevertheless, these sessionizing methods and heuristics have been shown to be highly accurate in session reconstruction. They therefore hardly affect subsequent data analysis measures (Spiliopoulou, et al., 2003). Standard heuristics are, for instance, the invariability of the IP address[28] within a session or an upper time limit between two page requests of 30 minutes. These premises of session reconstruction are not necessary if the data analysis is based on session-ids or cookies. Within our research, session-IDs were used in a number of our studies. Typical challenges of session reconstruction are cached pages that are not recorded

[28] Unless the data analysis is based on session-IDs or cookies.

in the data set and the use of proxy-servers/firewalls, which impede IP-based session reconstruction.

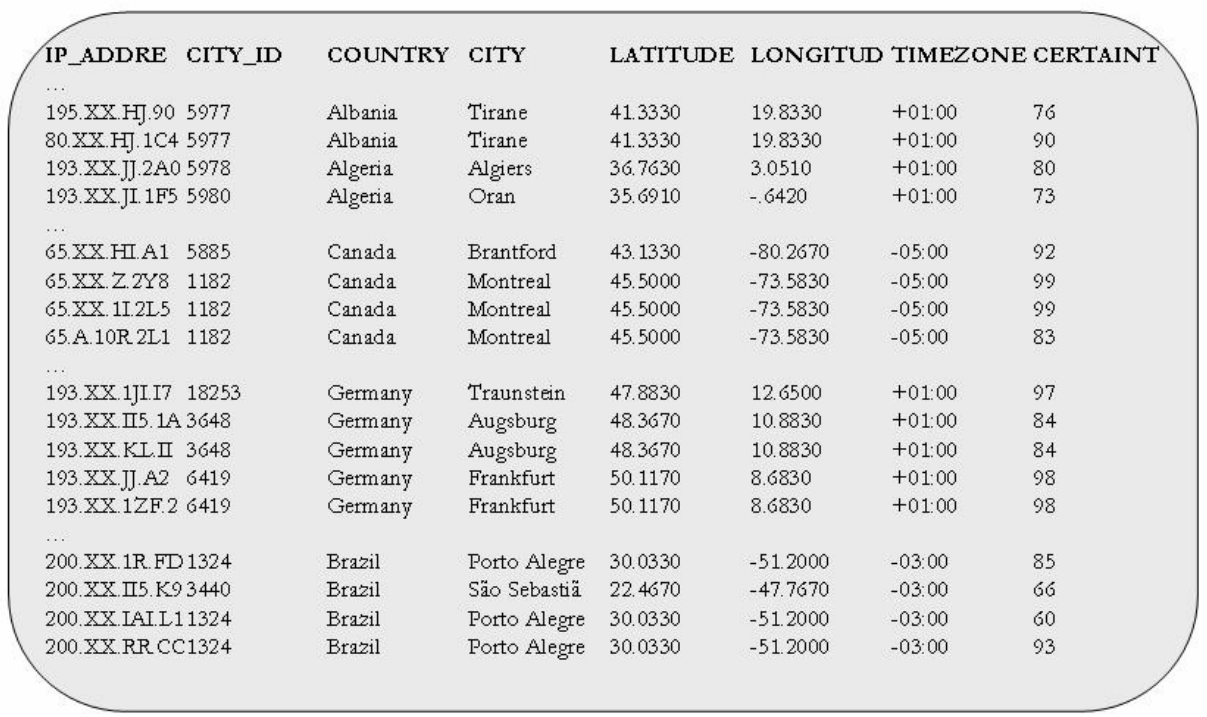
Logfiles do not only contain information about human user behaviour but also log page requests of robots. Robot detection is therefore, besides the filtering of the unused log entries (e.g. pictures are recorded separately from the requested page where they occur), an important part of data cleaning. Whereas so-called ethical robots may be easily detected and eliminated (e.g. by IP addresses), the detection of non-ethical robots is far more challenging and requires partial manual data cleaning. Tan and Kumar (2002) proposed a number of heuristics for detecting robots based on their navigation patterns (e.g. regularity of access, number of page requests) which were to a certain extent integrated into the used data preparation tools used.

The majority of the data preparation work was done automatically using the sessionizing tool WUMPREP, developed at Humboldt-University of Berlin and at Leipzig Graduate School of Management (www.hypknowsys.de). After data preparation, the data set was transferred into a MySQL database. We generated data subsets through appropriate queries. These subsets were eventually analysed with statistical software. Each time, a final data control was conducted when the logfile data was transferred to a database, based (for instance) on the number of page requests per session. Since long sessions are more characteristic for robots than for human beings, we set 100 page requests as a threshold and eliminated all sessions with a higher number of page requests. This was usually the case for only a few sessions per sample set.

A more comprehensive discussion of the problems of sessionizing and data cleaning can be found in Berendt et al. (2001).

For some of our studies, the logfiles were combined with a questionnaire posted on the same websites. In these cases we were able to skip parts of the data preparation process and were able to identify sessions thanks to the IP address or the session-ID.

Inferring Cultural and Linguistic Background. In order to also obtain information about the users' cultural and linguistic background in those the cases were information from a questionnaire was not available, we inferred the user's origin from the IP address. The employed GeoSelect software (www.geobytes.com) provides the following geographic information: country, province code, city, latitude, longitude, timezone and certainty of the information provided (Figure 5).



IP_ADDRE	CITY_ID	COUNTRY	CITY	LATITUDE	LONGITUD	TIMEZONE	CERTAINT
…							
195.XX.HJ.90	5977	Albania	Tirane	41.3330	19.8330	+01:00	76
80.XX.HJ.1C4	5977	Albania	Tirane	41.3330	19.8330	+01:00	90
193.XX.JJ.2A0	5978	Algeria	Algiers	36.7630	3.0510	+01:00	80
193.XX.JI.1F5	5980	Algeria	Oran	35.6910	-.6420	+01:00	73
…							
65.XX.HI.A1	5885	Canada	Brantford	43.1330	-80.2670	-05:00	92
65.XX.Z.2Y8	1182	Canada	Montreal	45.5000	-73.5830	-05:00	99
65.XX.1I.2L5	1182	Canada	Montreal	45.5000	-73.5830	-05:00	99
65.A.10R.2L1	1182	Canada	Montreal	45.5000	-73.5830	-05:00	83
…							
193.XX.1JI.I7	18253	Germany	Traunstein	47.8830	12.6500	+01:00	97
193.XX.II5.1A	3648	Germany	Augsburg	48.3670	10.8830	+01:00	84
193.XX.KL.II	3648	Germany	Augsburg	48.3670	10.8830	+01:00	84
193.XX.JJ.A2	6419	Germany	Frankfurt	50.1170	8.6830	+01:00	98
193.XX.1ZF.2	6419	Germany	Frankfurt	50.1170	8.6830	+01:00	98
…							
200.XX.1R.FD	1324	Brazil	Porto Alegre	30.0330	-51.2000	-03:00	85
200.XX.II5.K9	3440	Brazil	São Sebastiã	22.4670	-47.7670	-03:00	66
200.XX.IAI.L1	1324	Brazil	Porto Alegre	30.0330	-51.2000	-03:00	60
200.XX.RR.CC	1324	Brazil	Porto Alegre	30.0330	-51.2000	-03:00	93

Figure 5. Examples of Geographic Information Obtained from IP Address Analysis

Thanks to the detailed geographic data, we can attribute information about the users' native languages to each session, even in the cases of multilingual countries with different official regional languages (e.g. Switzerland, Canada). Language information about each region was collected in a self-generated language database that assigns one dominant language to each region or country.

We treated each (anonymous) user as a native speaker of the official, respectively dominant language in the geographic area from which s/he accessed the site. This choice is based on the following reasons: first, an official language is usually the native language of the majority of a region's inhabitants. The probability that the language chosen is the user's native language is thus higher than with other languages. Second, due to its official legal status, use of the official language is generally required for contacting public authorities and for other parts of everyday life (e.g., product labelling, description, etc.). It is therefore very likely that native speakers of other languages have at least a working knowledge of the official language where they live. In one case we did not apply the principle of the first official language, but opted for the "dominant" language (English instead of Irish in Ireland). In eleven cases we added widely spoken but not official languages (e.g., Russian in the former Russian Republics).

Each session was described only by the pages visited and the language group. Users were only categorized as L1 users if the webpages visited were presented in the user's native language. Hence, a

Spanish native speaker visiting the English version of a website would be classified as an L2 user. These cases were however rare.

It should be noted that the geographic information provided by the data was not specific enough to reveal information that might be in conflict with privacy laws or ethical issues. In the subsequent processing steps, IP addresses were removed completely.

Nevertheless, basing cultural and linguistic data exclusively on IP addresses and the inferred geographic information exposes some uncertainty in data validity. This is mainly due to the following limits:

(1) With some IP addresses, the software was not able to provide geographic information. These cases were identified by the software and excluded from further analysis. As illustrated in table 3, the non-recognized IP addresses represented up to 14 % of the sample.

Table 3. Percentages of Analysed IP Addresses through Geoselect Software

		Sessions	*Requests*
February	*IP address recognized*	117163	157196
	IP address not recognized	13414	16966
	Percentage of unrecognized IP addresses	11,45%	10,79%
March	*IP address recognized*	133723	178462
	IP address not recognized	16494	20893
	Percentage of unrecognized IP addresses	12,33%	11,71%
April	*IP address recognized*	82691	110484
	IP address not recognized	10973	13911
	Percentage of unrecognized IP addresses	13,27%	12,59%

(2) Even in the cases where the software inferred geographic information from the IP address, it is unlikely that the geographic information is correct in all cases. However, the software helps increase data validity through indicating the certainty to which the geographic information is correct.

(3) In addition, the logfile does not provide any information about whether or not the users accessed the Internet from their usual environment, whether their native tongue is

(one of) the official languages, whether they grew up with more than one native tongue and/or in more than one culture. Finally, a region's/country's language situation is not always easy to classify, so the language database can only reflect a simplified picture of the user's linguistic environment. Cross-validation of logfiles and questionnaires showed that the result of software-based native-language classification differed from self-reports in 6.7% of the cases (see also Kralisch and Berendt, 2005).

1.3.2.1.3 Additional Data from Web Crawling

Within our thesis, the main advantage of collecting data by means of web crawling is to complement data obtained from logfile analysis. A server log provides in the referrer data, information about webpages that are linked to the website. Yet, only websites that were visited by users prior to visiting our investigated website can hereby be identified as linked websites. Websites that are linked to the website but whose linkes were not used, were detected by a web crawler.. The web crawler developed for study 8 was in addition able to identify the language of the target webpage. This represents an advantage compared to logfile data, since another website's language can only be only inferred with caution from the top-level domain. Nevertheless, automatic language recognition carries a certain inaccuracy. It should also be noted that the web crawler used in study 8 obtained information from crawling search engines. Such an approach is based on the assumption that results from search engines are representative for the entire Web.

1.3.2.2 Questionnaires and Laboratory Studies

Data gathering through questionnaires can be considered a complementary tool to data recording through server logs. In contrast to logfiles, data from questionnaires can provide information about attitudinal variables, more reliable geographic and demographic information (e.g. the user's origin and residence native language, domain knowledge), and information about (other) control variables (e.g. Internet experience). Thus, questionnaires provided useful background information for analyses of logfile data. Accordingly, within our research, questionnaires are the main means of data collection in all studies that focus on the users' attitudes, beliefs and perceptions of language and culture related issues such as their satisfaction with the website (chapter 3, 4), risk perceptions (chapter 3), or aspects of perceived cultural and linguistic barriers (chapter 4). Questionnaires were furthermore used as a complementary means of data collection when data about the users' domain knowledge and other control variables were required in addition to the logfile information (chapter 2). Questionnaires also helped us to statistically validate the information we obtained from logfile data: geographical and language information inferred from the IP address were compared to the users' self-reported information. The fact that we were able to link each answer set from the

questionnaire to the respective logfile rendered this combination a rather comprehensive method of user analysis.

From a practical point of view, laboratory studies also have the advantage that participants can be asked a relatively high number of questions, which would normally lead to a high drop-out rate with online-questionnaires. This is useful when extended context information is desired (see below).

Internal validity is usually higher in questionnaires than in logfile analyses since information about additional variables can be assessed. These additional variables allow to control for potentially confounded variables (e.g. language proficiency level and domain knowledge) and thus to achieve a higher control of the impact of the investigated independent variable(s). With regard to the external validity, it can be stated that in comparison to logfile data, population validity is more restricted due to effects of self-selection. For example, among the website visitors the patient group was more represented than the doctor group, which can possibly be explained by more salient time restrictions with the doctor group. Self-selection is a particular challenge when users' attitudes are measured, because questionnaires are more likely to be answered by users that are with regard to certain aspects characterized by a similar basis attitude (e.g. more open to disclose data which might bias data regarding privacy attitudes). Ecological validity can be considered similar to that of logfile data if questions concern "tasks" that were carried out in the usual environment (e.g. the use of a particular website feature). However, the fact of filling out a questionnaire – i.e. the reactivity of the method – might bias responses, as well as might do changing perceptions if "task" and survey are separated through a longer time span. If situational variables (e.g. search goal) are considered to be a (potential) independent variable, internal validity is concerned. The internal validity is then lower in an online survey than in a laboratory study, due to the difficulty controlling of situational variables (e.g. Where and how is the website visited?)

Within our research we combined questionnaires and laboratory study in the cases where we expected a higher control of variables and thus a higher internal validity through laboratory studies. The control of additional variables allowed us to obtain supplementary information (e.g. the control of more in-depth behaviour analysis through the assessment of task-driven behaviour). Thanks to their higher internal validity, experiments are more appropriate for testing isolated cause-effect-relationships. External validity is first of all restricted due to the artificial laboratory setting which restricts considerably the ecological validity. It might be possible that laboratory studies augmented in our research one aspect of population validity by decreasing certain aspects of self-selection

biases. However, the fact that only students took part in the study as well as their limited number decreased other aspects of population validity. Furthermore laboratory studies often recruit the sample set from individuals who spent a certain amount of time outside their home country. In contrast, online surveys (or logfile data) can rely more on participants who live in their country of origin and are therefore more representative.

However, the evaluation of population validity is more complex. If significant cultural differences are obtained with participants that are less rooted in their home culture, it rather supports the findings of the impact of culture. In fact, it can be expected that cultural differences would be even stronger if tested with participants that are strongly integretated in their original cultural background (see also Hampton, 1979). The reason for the difficult evaluation of population validity within cross-cultural research lies in the fact that "culture" constitutes an independent variable. Culture as independent variable is inherent to the participants. Showing weaker cultural characteristics corresponds therefore to a weaker "treatment"[29]. Population validity in cross-cultural research is therefore more limited through other aspects. Students, for example, may differ from other people in their analytical skills or their valuation of financial aspects.

The same applies to the investigation of language. Language variables are also personal variables, i.e they are also inherent to the participants. From the fact that language and culture constitute personal variables results the difficult status of laboratory studies in cross-cultural and cross-linguistic studies. Language and culture as a characteristic of a person cannot be manipulated: (random) assignment is not possible. Accordingly, an examination of the effects of culture or language represents a correlative study. Due to the lack of possible manipulation of the independent variable, it is therefore very likely that that measured effects are due to other, correlating variables. Results are of correlative nature and should be interpretated only very cautiously – if at all – as a cause-effect relationship.

Nevertheless, the goal of our research consists first of all in investigating the effect of linguistic and cultural *correspondence* between the participants' characteristics and the website's characteristics. We compare for example how participants behave on an L1 website (i.e. high linguistic correspondence) or an L2 website (i.e. low linguistic website). Whether a participant is tested on an L1 website or L2 website can technically be assigned in a randomized way. However, due to restricted resources (i.e. the tested website is only available in a limited number of language

[29] Since "culture" is a personal variable, it constitutes a factor, not a treatment. The term "treatment" was chosen to show the analogy.

versions), *randomized* assignement is not feasible. Nevertheless, assignment and therefore manipulation of the independent variable is still possible. We argue therefore that such an analysis of language constitutes a quasi-experiment.

The same reasoning applies to the investigation of culture. Within our research, the difference between investigations of culture and investigations of languages consist in the fact that no culturally adapted versions of the website were available. Consequently, a treatment in terms of "cultural correspondence" was not possible in culture-related studies. The participant's cultural background is the only independent variable. Thus, as mentioned above, due to the lack of manipulation of such a personal variable, all studies are correlative studies. For a description and classification of experiments, quasi-experiments, and correlative studies see also (Campbell, et al., 2002; Sarris, 1992).

We employed two methodological tools to limit this risk: we assessed within the data collection potential correlating variables such as the quality of the health care system, and we aggregated cultural and linguistic groups so that each group was characterized by a certain variance with regard to background variables (e.g., in study 4 in Chapter 2 where the group of non-native speakers is represented by Russian and Malaysian native speakers – leading not only to different linguistic but also to diverging cultural backgrounds).

For the purpose of the thesis, two experiments were conducted. In the first case, the experiment's purpose was to obtain behavioural information concerning differences in the users' product valuation and in their attitudes towards data disclosure. Data was gathered as complementary information for data obtained from a questionnaire. The experiment was also the appropriate way to control for variables, which for ethical reasons, were not asked in the online survey, such as the users' financial resources, or use of paid online services (see section 3.3.2.3.2.2). In the second case, an experiment in task-solving form represented the only means of data collection about user preferences of structuring information (see section 2.3.5). More details are provided in chapters 2 and 3.

1.3.3 The Main Constructs

1.3.3.1 Conceptualization and Operationalization of Culture

Each of our cross-cultural studies employs a quantitative approach, based on Hofstede's index scores and hence on his cultural paradigm. Limitations linked to Hofstede's paradigm have been acknowledged in a myriad of articles and were illustrated in section 1.1.3. Yet, despite this criticism

we argued that Hofstede's model can be considered as appropriate for cross-cultural IS research if sufficiently but cautiously broadened. Depending on the study's objectives, Hofstede's concept of culture was extended by incorporating context-dependent approaches (knowledge level of culture) or by taking socio-economic variables into account. For the purpose of the thesis, two variants of a concept of culture are applied:

Basic Concept of Culture: Referring to Hofstede, culture is defined on the national level as the "collective programming of the mind which distinguishes the member of one group or category of people from another" (Hofstede, 1991, p.5). A cultural group is hence characterized by shared, yet distinctive, sets of values, beliefs, and thinking patterns resulting in behaviour and artefacts (see also Dahl, 2004).

Referring to Hofstede's cultural index scores, this basic concept of culture is furthermore characterized by the continuitiy of the "collective programming of the mind" over generations. It is therefore a stable concept.

A definition of cultural groups on the national level is first of all a consequence of referring to Hofstede's model. However, such a data aggregation is particularly useful if the market is approached on the same level. On the Internet, market segmentation by country might be less practical if access to a product/service was not restricted by language. Nevertheless, even online, product marketing is often segmented by country as a result of legal and logistical constraints.

Broadened Concept: The broadened concept of culture extends the basic concept by adding socio-economic variables and aspects of individual and knowledge-determined culture. We employ this concept in market-related studies that investigate individual needs of information and website features: Since these needs and preferences are very likely to be affected by current market alternatives and individual preferences, we consider the classic, stable approach as insufficient. For the purposes of our research we incorporated the quality of the Health Care System (as an example of socio-economic variables) and the users' medical knowledge and web experience (as aspects of culture that is based on an individual's knowledge). We argue that the context of a health-content website activates values and thinking patterns that relate to these three integrated aspects.

The quality of Health Care System was operationalized through the participants' ratings on a Likert scale. Depending on the investigation the number of questions asked differed. With respect to the quality of the Health Care System participants were asked to evaluate the overall quality or specific aspects (e.g. affordability, availability of doctors, etc.) of the system. With regard to web experience

participants were asked how frequently the used the Internet. Medical knowledge was measured by a dichotomous variable that distinguished between patients and medical professionals. Patients were treated as users with low domain knowledge and physicians as users with high domain knowledge. Details about the questions can be found in the studies 5 and 6 and the respective appendices (A-3.1, A-3.4, A-3.5).

Empirical comparative studies in cross-cultural research are usually conducted with two different national groups of opposite cultural values (Karahanna, et al., 2002). We however avoided this approach. In fact, a shortcoming of this research design is that measured effects cannot be clearly assigned to the impact of culture as defined by Hofstede. Since culture is defined on the national level in empirical studies, the risk that effects were caused by correlating variables is high (e.g. income effect). In studies with a broadened concept, we control for a few of these variables. However, we cannot ensure that other, potentially correlating variables do not have an impact. As an alternative, we analysed sample groups composed of individuals from different countries. Depending on the research objectives, the impact of culture was then measured in one of the two following ways. We either calculated correlations with cultural values across all individuals, or we assigned individuals into two groups according to their cultural values, resulting in a comparison of the test groups. The boundary between the low and high groups equals the highest cultural index score divided by two (see also Hofstede, 2001).

A multi-nation approach is also useful from a practical point of view. The sample set - if obtained from online surveys - does not, in general, provide sufficient data from comparing specific countries. Further details of the operationalization of culture are illustrated in the respective sections.

1.3.3.2 Language

1.3.3.2.1 Conceptualization and Operationalization of Language

Language is, in our studies, predominantly examined from what we called the "user-perspective" (vs. "service provider perspective"), i.e. regarding the users' cognitive effort when visiting L2 websites and their attitudes towards them. However, the insight revealed is also helpful when inferring recommendations for website providers. In all languages studies, website visitors are classified as either native (L1) or non-native speakers (L2) with respect to the languages offered on the website. Similarly to our way of measuring culture, this approach allows the control of, to a large extent, the impact of culture, since each language group is likely to exhibit a large variance in cultural backgrounds. We also assessed non-native speakers' proficiency levels, mainly for

descriptive but also for analytical purposes, if language data was gathered by means of a questionnaire. In one survey-based study domain knowledge represented a second dimension of language proficiency (for reasons see section 2.3.4). In section 2.3.5, we focus on structural aspects of language and therefore compare native speakers of four different languages: German, Russian, Malay, and English. Attitudinal language variables focus on individuals' perceptions of the Internet language offer and their attitudes towards L2 websites (see section 4.3.2).

Study 9 in chapter 4 emphasized the market/provider perspective of language. Within this study we compared individuals' *perceptions* of the amount of native language information available on the Internet with the actual amount of native languages online as specified by Internet statistics.

1.3.3.2.2 Limitations of our Approach

A first drawback of our studies consists of the simplified assessment of users' language skills. Sophisticated neuro- and psycholinguistic examinations allow for a concise identification of native and non-native speakers, as well as specific linguistic tests of foreign language proficiency levels. The complexity of these examinations and the extent of these tests were however not deemed to be justified for the purposes of our research. Assignments to one or the other group or to a certain proficiency level were therefore determined in survey-based studies of users' self-ratings. We furthermore restricted the gathering of data about foreign language proficiency to language skills in English.

The limitations of conceptualizing and measuring language attitude (as any emotions in psychological measuring) are well-known (e.g., Dawes, 1972) and apply also to our research design. In particular, a clear distinction between the role of language on the one hand and the role of language-associated values is difficult to obtain. However, it was not deemed necessary for our research objectives and therefore not assessed.

Finally, in our comparisons of native and non-native speakers we did not distinguish between different levels of literacy attainment in L1 languages. It is often argued that literacy attainment is an important determinant of information consumption on the Internet (e.g., Sims, et al., 2005). Nevertheless, such an additional differentiation goes beyond the purpose of this thesis, and would overburden analyses due to increased complexity.

1.3.4 Objects of Investigation

Two websites provided the necessary database for collecting data for our studies. Both websites offer health information. For reasons of confidentiality we refer to them as website A and website B.

1.3.4.1 Website A

All of our studies but one were conducted using website A: we analysed the website's logfiles, posted questionnaires on the website, and also based our laboratory experiments on it. The website offers information about skin disorders and is visited by medical professionals as well as by patients worldwide. Information on the website was, until 2002 available in four languages (English, German, Spanish, and Portuguese). French was added as an additional language in 2003. All of the website's language versions were presented with the same interface design. Figure 6 depicts a simplified view of the website's core structure.

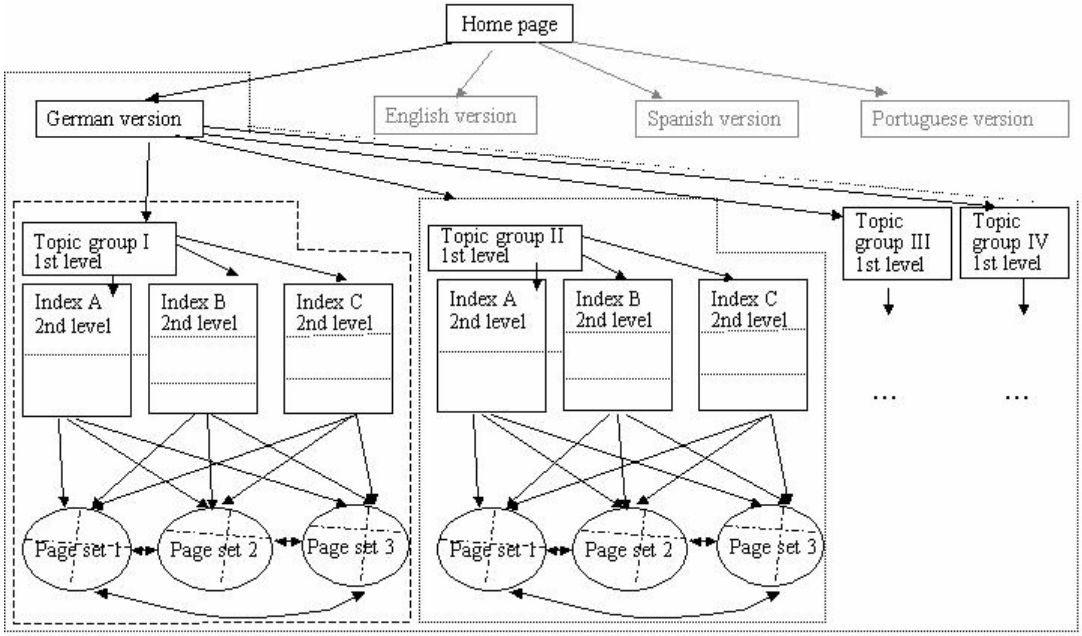


Figure 6. Simplified Website Structure
1st level = entry page for the concerned topic group (here: group of diagnoses)
2nd level = various search modi / browsing options
Page set = collection of individual pages regarding the same sub-topic (here: diagnosis)

Details concerning available search options and navigational structures are presented in chapter 2.

1.3.4.2 Website B

The use of website B is restricted to medical professionals. In contrast to website A it offers an interactive service: website users can ask and answer medical questions in the website's discussion forum and private consultation services. Similar to website A, users of website B have different linguistic and cultural backgrounds. Website users on website B almost exclusively use English in their communication.

1.3.4.3 Participants

In our research we essentially referred to two major types of sample groups: website visitors and university students. Logfile analyses and questionnaires assessed data about the visitors of website A and website B. This type of sample group encompasses medical professionals and patients (not on website B) worldwide.

Experiments were conducted at a Humboldt-University Berlin and at a major Malay university involving German students, international students studying in Germany, and Malay students of different ethnicities (Malaii, Chinese, Iban).

With regard to the first type of sample group, a bias is first of all caused by Internet distribution in the different parts of the world as well as by telecommunication costs. Since we did not include non-Internet users in our studies, it can be assumed that less developed countries are overrepresented by physicians who are more likely to access the Internet.

Difficult access and high telecommunication costs are also likely to lead to an overrepresentation of Western countries in the survey sample. Further bias-inducing facts of self-selection are the languages of the surveys and time required by filling out the questionnaire. The language(s) chosen for the survey encourage primarily native speakers and non-native speakers with a high proficiency level to participate in the survey. Differences in proficiency levels between native speakers and non-native speakers and resulting outcomes can therefore be assumed to be more pronounced than results from the survey data. Furthermore, it can be expected that physicians are underrepresented in the questionnaire sample group since they usually have less time available to answer questions online.

The appropriateness of relying on students in general for conducting experiments, and those who have already spent a certain amount of time in their guest countries in particular, was already discussed in section 1.3.3.1 when we presented arguments in favour of this sample group.

Further discussion of the problems of cross-cultural research with regard to possible bias and inequivalence of constructs, methods, items, or in sampling can be found in (Brislin, 1986; Evaristo, 2004; Evers and Day, 1997; Hui and Triandis, 1985; Karahanna, et al., 2002; Mullen, 1995).

1.3.4.4 Health Websites as an Application Area

Within our research we investigated the role of language and culture on websites that offer public health information. This type of website was appropriate for the purposes of our analyses for a number of reasons.

As mentioned in the introduction, health websites usually pursue societal goals, but due to increasing financial restrictions, also work under commercial constraints. Therefore, the application of our research into the field of health education covers these two types of websites.

Health education has been gaining importance over the last few years. In many cases, the Internet offers a number of advantages compared to the offline world, such as when meeting a physician involves high costs or elicits strong feelings of shame. For the information provider, the Internet is a medium of relatively low cost, and is therefore appropriate as a central medium of mass health education. As a consequence of increasing amounts of public information transmitted over the Internet – including those of health education - demands for equal access to online information are growing in importance. It is becoming essential to enable all citizens to fully participate online, regardless of their age, gender, education, or other personal characteristics. In fact, current initiatives often see the success of the new methods of education as too optimistic given the difficulties that the medium generates for some parts of society (Sims, et al., 2005).

Language and cultures are two of the characteristics that potentially impede access to (health) information. In particular, cultural but especially linguistic groups that are so far underrepresented on the Web are subject to these inequalities. A health website that offers several languages allows us to compare the effect of native and non-native language information offerings, but also that of cultural adaptations with regard to such an important topic as health education.

Health websites furthermore provide the advantage that they also easily allow the comparison of users of low domain knowledge (patients) and high domain knowledge (physicians). This aspect is particularly important with respect to linguistic analyses.

The appropriateness of health websites is also justified through the fact that health content is, to a certain extent, equally important worldwide. The basic relevance of the website's content should

consequently not give rise to differences in user behaviour. This constitutes an important assumption for a number of our empirical studies, in particular those that analyse the role of culture.

Websites offering public health information first of all pursue an educational goal. With respect to this goal, the impact of language and culture is predominantly examined from a cognitive point of view, investigating access to information and their ease of understanding (Chapter 2). Chapter 4 provides further insight by investigating the role of language as a barrier to distributing and accessing information.

Health content websites are also subject to a number of social and attitudinal issues that affect the perception of the health information as a customer service. These issues pertain to the relevance and need of specific types of information depending on the cultural background, leading to differences in risk perception and perceived benefits from the website's features and services (Chapter 3). The role of language as a determinant of website satisfaction is examined in chapter 4.

Whereas cultural and linguistic differences in the categorization of diseases are specific to health content websites, it can be assumed that the vast majority of the rest of our results can be applied to websites with diverging content and services. Future research is however needed.

1.3.5 Units of Analysis

In the following three chapters we present nine studies. It should be noted that "study" refers here to content-oriented units of data analysis rather than units of data collection. If within one study data came from different sources, we split the analysis into sub-parts, as indicated by letters (a, b, etc.).

In a few cases, the same questionnaire or the same logfile material was used to assess data for two different topic areas (e.g., study 6 and 7). Due to the independence of the topic areas and the independent examination of each hypothesis we did not correct the alpha-significance threshold through Bonferroni correction (see also Bortz, 2004, p. 129-130).

1.3.6 Overview of Measures and Statistical Procedures

Within the empirical sections of this thesis a number of statistical analyses are used, differing in terms of procedures and (categorization of) variables. This section aims to provide an overview of the various analyses employed.

In our studies, three major forms of statistical procedures are applied. These are: analyses of correlations, analyses of variances, and regression analyses. Selection of the appropriate statistical procedure is based on the goal of the study as well as statistical requirements (e.g. metric vs. non-metric). With regard to the study's goal, our hypotheses focus on either identifying the *relationship* between cultural variables and the dependent variables (correlational analyses), or on establishing *difference*s between (two) culturally or linguistically different groups (analyses of variance), or on determining *how* culture and language affect behaviour and attitudes (regression analyses) (see also Bortz, 2004, p. 181, 203, .207, 243; Cohen, 1968).

Analyses of correlations examine the relationship between variables, without proving causation. This type of statistical analysis seems to be most common in cross-cultural analyses (e.g., De Mooij, 2003). In cross-cultural studies as well as other socio-anthropological studies, variables seem to be determined by a considerable number of factors. Proving causality appears to be particularly difficult since it would require the elimination all non-cultural factors (White and Korotayev, 2004). Yet, this is not trivial. This thesis addresses the issues of culture as a variable that is inherent to the test participants and the challenging methodological implications several times (e.g., section 1.3.3.1). Due to this link a clear separation of culture from other variables is not always possible. It is therefore particularly difficult to provide evidence for causal effects. Hence, in cases where a simple identification of the relationship between culture and the dependent variables investigated was considered sufficient, we applied correlational analyses. In the case of interval variables, Pearson's product-moment correlation was used (study 1); in the case of ordinal variables, Spearman rank correlations (study 6, 7, 9). We kept the original interval scale categorization as proposed by Hofstede's cultural index scores when investigating the stable concept of culture.

In other studies we focus our hypotheses on providing evidence for existing differences between groups. This applies in particular to studies where we examine the impact of language. Language is treated as a dichotomous factor for the most part, distinguishing between native speakers and non-native speakers. In these cases we carry out analyses of variance (ANOVA). Depending on the number of independent variables, analyses of variances are one-way ANOVAs or k-way-ANOVAs. In study 2, 3 and 6 multivariate ANOVA (MANOVA) is applied, as a result of existing correlations between the dependent variables (Bortz, 2004, p. 586). MANOVA results are based on Pillai's trace, due to the robustness of this test (Olsen, 1979). Culture (stable concept) was categorized as a factor with two values (high and low) here, transforming Hofstede's cultural index scores into dichotomous variables with a cut-off point of 50 (see also Hofstede, 2001 and section 3.3.2.2.3).

In study 4 the dependent variables are nominal variables. Consequently, group differences were tested through the Kruskall-Wallis-H-test, as the non-metric equivalent of the ANOVA procedure (Bortz, 2004, p.287).

Regression analyses were applied in studies where we investigated *how* culture affects behavioural and attitudinal variables. In those studies we examined the impact of factors other than culture as well as their relationship between each other at the same time. Regressional analyses can be found in study 5 and 9.

We also used additional statistical procedures for data preparation and pre-analyses. Normal distribution was tested by means of the Kolmogorov-Smirnov-test. Chi-square tests examine the distribution of variables within each of two tested groups and test so the independence between a dependent and an independent variable. Due to their low statistical requirements (Bortz, 2004, p. 154) chi square tests were predominantly used to test whether two investigated groups differed with regard to a control variable (study 6, 7). In study 3b, the Mann-Whitney-U-test was applied for this purpose.

In study 4, 6, 7, 9 attitudinal measures are assessed through several items. We tested the items' appropriateness using reliability tests (Cronbach's alpha). Factor analyses were applied to condense data by reducing the items into factors (Bortz, 2004, p.513). In a few studies, if reliability was sufficient, the statistical mean was used to aggregate data (study 5, 9). Factor analyses were always carried out as principal component analyses with varimax rotation (see also Bortz, 2004, p. 548-553) and eigenvalues of 1. Only in study 7 was the eigenvalue criterion relaxed to 0.9.

Finally, due to the nature of the variables language and culture, sample sizes per group usually differed. In order to reflect the different sample sizes, we applied weighting procedures, using the number of participants per group as the weighting measure (study 3, 4, 7) (Bortz, 2004, p. 39).

2. CHAPTER:

THE IMPACT OF CULTURE AND LANGUAGE ON WEBSITE DESIGN PREFERENCES: NAVIGATIONAL STRUCTURES, SEARCH OPTIONS, INFORMATION CATEGORIZATION

2.1 INTRODUCTION

Our empirical studies are presented in three chapters. Chapter 2 includes all studies that investigate the impact of language and culture on preferences for information *presentation*. Studies regarding culturally determined preferences for information *content* are subsumed in chapter 3. Studies in chapter 4 emphasize market aspects; they investigate for example how the number of speakers affects the language offer on the Internet, how, as a result, language may form a barrier to accessing information, and how the total language offer online affects the satisfaction with the language offer of a particular website. The order of the chapters follows more or less the categorization that was introduced when the role of language was discussed: chapter 2 and 3 focus on the user-perspective; emphasizing in chapter 2 mental models and cognitive-behavioural aspects. Chapter 3 concentrates in contrast on attitudinal issues. The role of the market, key issue of the provider-perspective, is subject to the empirical investigations in chapter 4.

This chapter presents a collection of studies investigating culturally and linguistically determined preferences for website design. It examines language's effect on cognitive effort vis-à-vis various search options and forms of information categorization. Preferences for search options and navigational preferences are also investigated from a cultural perspective. Based on the findings from our studies we infer implications for website design at the end of this chapter.

These studies and their goals share a number of commonalities with usability studies. The goal of usability research is to adapt a website's functionality and content to its users' preferences in order to facilitate use, and consequently the adoption of services and information offers. The term *usability research* appears generally refers to a specific product that needs to be tested within one or more specific target groups. We also employ the term *usability research* to describe broader analyses that aim to gain more widely applicable and generic knowledge about the impact of language and culture on the needs and preferences for certain website design features and structures.

Insight gained from usability analyses is often issued in the form of guidelines and directives containing heuristics on how to adapt user interfaces to the users' specific needs and preferences,

and how to evaluate localized websites (Mandl, 2005). An early comprehensive and important contribution towards cross-cultural adaptation was provided by the guidelines published in 1996 by Nielsen and DelGaldo (1996). Nevertheless, as Mandl (2005) remarks, guidelines only usually regard rather specific and tangible issues such as country-specific formats, colours, and symbols (e.g., Nielsen, 1993; Romberg, et al., 1999). Intangible aspects of cross-cultural differences such as knowledge organization are mostly disregarded. It seems that the impact of intangible aspects has been primarily described in single scientific papers.

2.2 LITERATURE REVIEW: COGNITIVE LOAD, THINKING PATTERNS AND OBJECTIVE ACCEPTANCE OF INFORMATION SYSTEMS

Aspects of usability have traditionally been investigated by scientists from various disciplines. With regard to the cognitive load and *"objective usability"* (Ford, et al., 2005) of Information Systems, psychological studies have provided the most valuable insight. Yet, these studies refer to individuals as "a single mind interacting with an isolated technology in a social, cultural, and historical vacuum" (DePaula, 2003). Kamppuri and Tukiainen point out that "even though anthropology was originally seen as one of the fields that should form cognitive science, its role in cognitive science has been very modest to date" (Kamppuri and Tukiainen, 2004).

Cross-cultural usability studies, however, focussed for a long time on the assessment of differences regarding visible aspects such as colours and symbols. They were followed by various studies where attitudes towards these features and also their understanding were gauged. Analyses that examine the impact of culture on behavioural aspects are however still rare. Also, less visible features have become the subject of cross-cultural usability research only recently.

An important contribution was provided by Marcus and West Gould (2000) who showed how a website designer's cultural background affects the design of a website. Their study goes beyond an analysis of the outer layers of culture (such as the use of symbols) and also examines the other features through which intangible cultural core values are expressed on a website. As such, the authors oppose examples of typical website design, expressing for instance Masculinity and Femininity, high Power Distance and low Power Distance, and Long-term Orientation and Short-term Orientation. In contrast to earlier cross-cultural usability studies, they investigate differences in the website's structure such as its symmetry or hierarchical levels or arrangement of hyperlinks. Even though this study has been criticized for its methodological approach and low validity, it however represents the basis for a number of further studies in this field (e.g., Cyr, et al., ; Perugini, et al., 2004; Sun, 2003).

Studies that are presented in this chapter can be understood as a replication of Marcus's and West Gould's analyses: whereas Marcus and his colleague focus on the creative process, our studies emphasize cross-cultural differences in *usage* patterns aiming in parts to mirror Marcus et al. findings.

Language is a rather rare subject in studies of international usability. Within academic IS research, studies analysing culture seem to dominate those that analyse language. This miscorrelation might be due to the fact that language is often considered a sub-category of culture in usability research. It therefore draws less attention as an independent subject. In most studies language is considered as one – overt - aspect of culture. Even though language barriers on the Internet are much more visible than cultural barriers, the predominant focus on language only occurs in the field of Information Systems in selected investigations. Being rooted in marketing research, language-focussed research is mainly investigated with regard to its effect on attitudinal variables in international usability.

Luna and Peracchio (2002; 2003) studied the effects of language processing and cultural patterns and preferences, their interaction, and their impact on website evaluation. With their empirical analysis the authors demonstrate that cultural congruity between cultural manifestations on the website (e.g., design elements, content) and the visitor's cultural values may moderate a language's effect on attitudinal variables, such as persuasion.

Sociolinguistic studies look at how languages change as a result of the impact of the Internet (e.g., Evans, et al., 2004; Pargman and Palme, 2004), and the need for languages other than English (in particular minority languages) on the Internet in general and for e-learning in particular (Herring and Estrada, 2004; Keegan, et al., 2004; Stander and van Belle, 2004 - see also section 4.3.2).

However, language's impact on various behavioural aspects of website use, in particular with a cognitive focus, has hardly been considered within the context of Information Systems.

As mentioned previously, the use of online health websites depends on how easily users can access the information sought, to a large extent. This chapter therefore outlines the link between search option use and the users' cultural backgrounds and language skills. Accordingly, the studies in this chapter focus on how users prefer to search for information on a given website, dependent on their cultural and linguistic backgrounds.

2.2.1 Navigational Structures

Differences in navigational behaviour were originally analysed in psychological studies. Navigational patterns were examined with regard to two aspects: first, the extent to which they are influenced by a website's structure, and second, how they are affected by individual differences in cognitive styles (Berry, et al., 2002). Psychologists have studied the influence of menu depth and breadth on search time, search efficiency, degree of disorientation, and satisfaction (e.g., Jackson, 1997; Larson and Czerwinski, 1998; Snowberry, et al., 1983). Results from these studies provide evidence that a website's structure has a considerable effect on navigational performance. Adapting websites leads to better navigational performance and easier information access. Otter & Johnson (2000) find, for instance, that hyperlink structures that are predominantly based on associative links[30] are more likely to cause a feeling of being lost. Positive correlations between navigational performance and user satisfaction have been identified (Otter and Johnson, 2000; Smith, 1996). Other studies show how individual differences in structuring and processing information affect the way a user navigates on a website (e.g., Chen and Rada, 1996; Dufresne and Turcotte, 1997; Swan and Allan, 1998) Results from Dufresne et al. (1997) suggest that so-called field independent[31] users usually explore a website to a larger extent than field dependent[32] users do. Cultural aspects were however not integrated into these psychological studies.

A major contribution towards classifying human navigation behaviour was provided by McEneany 2001 (2001). Based on works about websites' network structures (Botafogo, et al., 1992), McEneany developed metrics that describe human navigation paths through these hyperlink structures. Due to the complexity of the calculations, it was not possible to apply these metrics to the large amount of data we analysed in our studies.

Insight into differences between single individuals is, in a certain sense, of limited value since adaptations of a website's navigational structure can hardly be personally customized. Taking results from cross-cultural research into account can therefore be useful, if culture is understood in terms

---

[30] "Associative links rely exclusively on source-destination content similarity …" ("see-also…-link"). They are opposed to structural links, which are made explicitly within the site's hierarchy. Danielson, D.R. (2002): Transitional Volatility in Web Navigation: Usability Metrics and User Behavior., M.S. Thesis, Symbolic Systems Program, Stanford University. URL: http://www.stanford.edu/~davidd/MastersThesis/Danielson-Transitional_Volatility_in_Web_Navigation.pdf

[31] "Field Independent people tend to have good analytical and cognitive restructuring skills. They will actively reorganize information according to contextual demands and impose structure when necessary according to their experience. They are likely to form a mental model of the situation before proceeding with their task." (Dufresne et al. 1997)

[32] "Field Dependent people tend to adopt a passive approach in learning and problem solving. They prefer to be guided and to rely on external referents. Perception is dominated by the prevailing field. When internal referents are less available, FD people are more likely to respond to the dominant properties of the field as given." (Dufresne et al. 1997)

of the distribution of certain cognitive styles, needs and preferences of the population of a country (concept 1). Since cognitive preferences correlate with navigational performance and website rating, users' cultural backgrounds are expected to do the same.

From a cross-cultural perspective, insight has been provided solely with regard to website designers' preferences for structuring websites. Marcus and West Gould (2000) identify a preference for symmetric design in high power distant cultures, or for a "clean, functional" design allowing fast information access in short-term oriented cultures. Furthermore, low uncertainty avoidant cultures, in contrast to high uncertainty avoidant cultures, feature a complexity of content and large navigational choice. However, little insight has yet been provided about cross-cultural differences in navigational behaviour.

2.2.2 Search Options

The study of information retrieval systems has a long history. A shift in the focus of attention towards the user of the system only occurred in the 1970s (Case, 2002). Whereas an analysis of navigational structures and patterns gives an overview of website use, a focus on the use of search options goes into more detail regarding one particular aspect of website use: the search process. Nevertheless, a clear distinction between navigational and search behaviour is not possible: on the one hand navigating on a website is often done for the purpose of finding information; on the other, seeking information usually involves navigation on the website. As long as no information about the user's goals when visiting the website are available (such as in our case), the difference is mostly on the level of analysis. Our study of navigational behaviour focussed predominantly on how given hyperlinks are followed, emphasizing the depth and breadth of the hyperlink structure. This analysis of search behaviour emphasizes which kind of hyperlinks one follows (e.g. alphabetical or content search). The further incorporation of search engines are included in the analysis.

Nevertheless, the interrelationship between culture and cognitive styles and hence the role of culture are expected to rely on the same principles. Besides the impact of culture, the impact of language on information retrieval is investigated.

Human information seeking behaviour has been studied from various points of view. The literature offers a range of classifications of search behaviour and its determinants. One common kind of classification is based on the objectives of information seeking; it differentiates between goal-oriented and exploratory search modes (Dholakia and Bagozzi, 2001; Hoffmann and Novak, 1996;

Moe, 2003; Nielsen, 1997). Another influential kind of classification is based on the *process* of information seeking (see Rice, et al., 2001 for an overview). Marchionini (1995) distinguished between search strategies and browsing strategies. Analytical *search* strategies are, to a large extent, planned, goal-driven, and deterministic. Classic information retrieval using queries is an example of this type of search strategy. In contrast, browsing strategies are of a rather heuristic, interactive, and opportunistic nature. Browsing strategies as a form of information seeking have primarily emerged with the rise of hypermedia and hyperlinked text forms. Although often used as a synonym for browsing, in Marchionini's framework navigation represents only one of four browsing techniques (beside scanning, observing, and monitoring), (see also Cunliffe, et al., 1997). Nielsen (1997) distinguished between search-dominant and link-dominant behaviour and empirically measured the distribution of these search types among users.

Marchionini (Marchionini, 1995; Marchionini, et al., 1990), emphasized that individual characteristics and experiences determine how and what kind of information is extracted. Search expertise as a determinant of information seeking behaviour has also been investigated by other authors including Ingwersen (1982) and Hölscher and Strube (2000). Further personal characteristics influencing search patterns include education (Ellis, 1989) and uncertainty (Kuhlthau, 1993). Navarro-Prieto et al. showed that experienced web users preferred the use of search engines whereas novices seemed to be inefficient with their searches. The authors also found evidence supporting the impact of topic knowledge. The role of topic knowledge is discussed in more detail in Section 2.2, together with the role of language as another aspect of cognitive capacity. Similar studies were conducted by (Vakkari, 2000).

Taylor (1962) identified four different levels of information need: visceral (perception of need), conscious (mentally defined), formalized, and compromised (mapping natural language into an adequate search system language).

Building upon these personal and situational variables, Wilson introduces the influence of demographic (e.g. gender, social and economic status, education, job experience) and environmental variables (legislation, economic situation, information culture) to his model of information seeking. Investigations into the impact of culture as an umbrella construct are however a very recent subject with little insight so far. In 2001 Iivonnen and White compared initial search strategies of Finnish and American students confronted with various search tasks. The resulting differences between the groups suggest an impact of the users' cultural backgrounds. The study was however of a rather exploratory character, since the group characteristics found had a weak

foundation in cross-cultural theory. Other studies focussed again on the visible features of graphic user interfaces such as colours and icons (e.g., Duncker, 2002). A rather frequent subject is the study of information-seeking behaviour of ethnic minorities. Their usefulness for the purpose of our research is limited since "they do no systematically compare variations in search behaviour" and "are also […] focussed on questions related to library service" (see Komlodi, 2005). Following our own research contribution (Kralisch and Berendt, 2004), Komlodi and Carlin (2004) summarize potential impacts of culture (as expressed in Hall's and Hofstede's cultural dimension) on information seeking behaviour in a systematic manner. Komlodi expands the theoretical contribution to the subject by identifying fundamental elements for a systematic study of cross-cultural differences in information seeking behaviour. Nevertheless, academic literature still largely lacks empirical support for the theoretical bases.

With respect to language, the focus on the system rather than on the user is still strong. System-based multilingual information retrieval challenges scholars and practitioners with respect to the creation of a multilingual corpus, the query formulation process and system evaluation. A more detailed illustration of the topic can be found in (Mandl and Womser-Hacker, 2004).

Given the significance of individual cognitive differences as shown by previous information retrieval research, language is likely to affect users' search behaviour too. In a study of information search using textbooks (Goldman and Duran, 1988), non-native English speakers tended to spend more effort and more time understanding a search target passage than native speakers, indicating that the same task involved different cognitive effort for the two groups.

Many aspects of human behaviour obey the law of efficiency. With regards to Internet navigation behaviour, the Information Foraging Theory (Pirolli and Card, 1995; Pirolli and Card, 1999),(Pirolli and Card, 1999) proposes to explain user strategies for seeking, gathering, and consuming information on the Internet as a result of the effort involved. Language skills and domain knowledge should consequently influence search behaviour on the World Wide Web as well as within a website. Mediation effects between the two variables can also be expected. It is therefore helpful to understand the underlying cognitive mechanisms and interdependencies as illustrated in the following sections.

Nonetheless, search options are characterized more by dimensions than their linguistic features. The search options differ, for example, in how fast search results can be obtained or how many clicks are necessary to find the information sought. Thus the decision about which search option to

use is a trade-off between linguistic requirements and cultural habits, and other characteristics of the search option that might be useful for reaching the search goal.

2.2.3 Information Categorization

This category represents a further point of view on analysing information seeking behaviour and their consequences on a website's design. It is strongly intertwined with the two categories above, since a website's hyperlink structure and search options inherently encompass a categorization of the information offered. For instance, the international medical catalogue WHO ICD classifies diseases by their causes and thereby creates a structure of information classification. Another – very common – way of categorizing information is an alphabetical index, which emphasizes, for example, verbal cues instead of visual cues.

The classification of information strongly focusses on the content and therefore on the knowledge of the topic. A myriad of studies underline the role of domain expertise. A few of them were mentioned above. Koch (2003) found that browsing through content-based classification schemes is especially useful to users who are unfamiliar with the subject domain and its structure. One of the advantages of this type classification scheme is that they provide a context to guide the user's search of information access.

Nevertheless, Information Systems literature provides little insight into the impact of language and culture on preferences of information classification. Research suffers furthermore from the fact that most studies are specific to one culture and can hardly be applied in wider contexts. Blake (2001) found that animal experts in a South African National park who were functional illiterates did not want to have hierarchized information within the information system. Such a decline of hierarchical structures would be rather uncommon in Western cultures. Yeo and Loo (2004) identified differences in preferred food categorization between the various ethnicities of Malaysia.

With the specific respect to language, studies were conducted in the field of psycholinguistics and mainly regard aspects of the Sapir-Whorf-Hypothesis as presented in section 2.3.5.1. Differences in information categorization between native speakers and non-native speakers have not been examined yet.

2.3 EMPIRICAL WORK

2.3.1 Overview of Participants, Material, and Procedures in Study 1 to 3

Study 1, 2, and 3 are based on the same logfile data obtained from website A. Their empirical analyses are therefore similar. Commonalities are presented below (section 2.3.1.1 to 2.3.1.2), whereas information specific to each analysis is portrayed in its respective section.

The study of information categorization preferences (study 4) differs somewhat from the first three studies. In addition to behavioural data it also assesses attitudinal data. Consequently, since a higher control over situational variables (especially search tasks) was necessary, data was gathered in a laboratory experiment. Details of the method of study 4 are therefore described separately, in section 2.3.5.

2.3.1.1 Participants

The navigation behaviour of users from 188 countries including 9,971 different cities is documented in the logfile data obtained from website A. The questionnaire for study 3 was answered by 165 users originating from 34 countries, speaking 15 different native languages.

2.3.1.2 Materials and Apparatus

For the study of navigation and information seeking behaviour (study 1 to 3), data was gathered from the usage log of website A between November 2001 and November 2002. For a simplified illustration of the website's structure, please refer to section 1.3.4.1.

Sessions that describe each user's navigation path through the website and their use of search options are inferred from the logfile by means of the procedures illustrated in section 1.3.2.1.1.

Logfile cleaning followed current standards as described in the same section. We employed the sessionizing tool WUMPREP and complemented the cleaning procedure by a manual quality assurance step. For easier data processing we transferred the logfile data to a database. Final data calculation was carried out using statistical software.

Website A offers the following search options: search engine, alphabetical search and content-organized hyperlinks. They differ in how directly and quickly information can be accessed, how much context information is provided, how many hierarchy levels need to be traversed, and regard the required vocabulary knowledge.

The following two figures provide an overview of the search options investigated. Figure 7 displays the taxonomy applied in cultural studies, encompassing dimensions of time, hierarchies, and

context information. The taxonomy presented in Figure 8 is used for investigations of the impact of language. It contains the dimensions of language proficiency and domain knowledge.

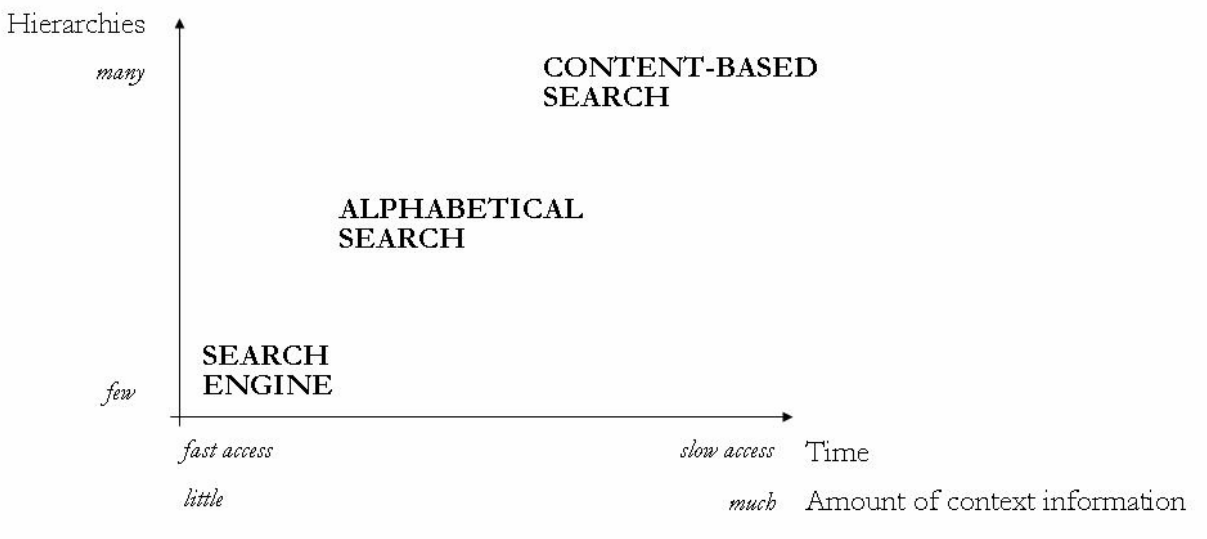


Figure 7. Taxonomy of Search Options for Cultural Studies

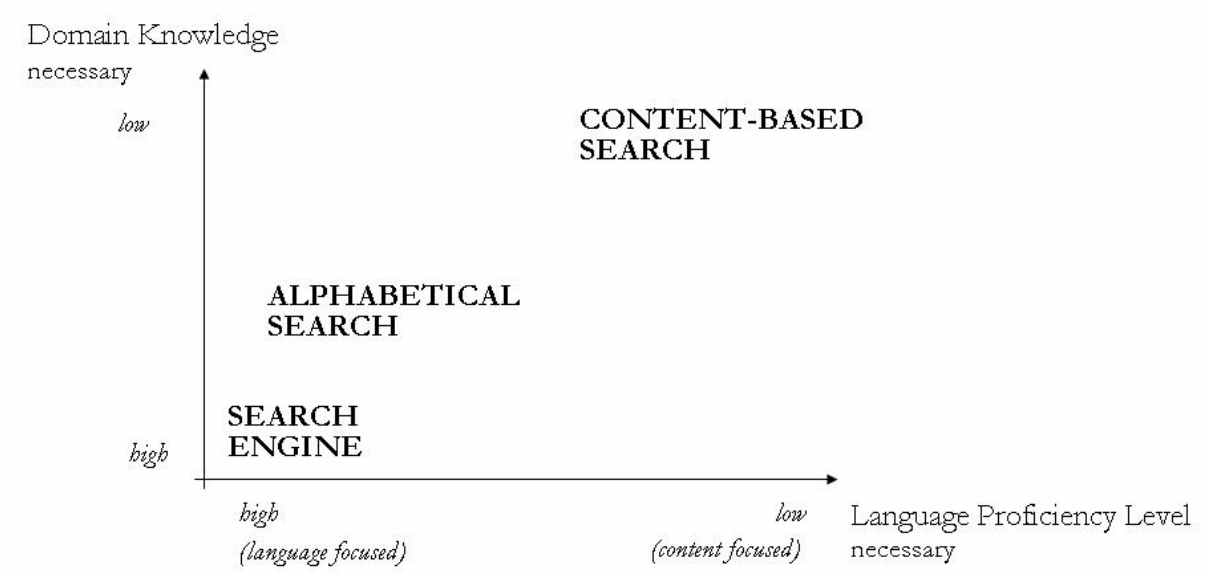


Figure 8. Taxonomy of Search Options for Linguistic Studies

Details concerning these taxonomies can be found in the sections 2.3.3.1.1 (culture) and 2.3.4.1.1 (language).

The vast majority of the information necessary for our analysis is provided by the logfile (e.g., access of search engine, hyperlinks, language of webpage etc.). Additional information about the users' medical background and other personal variables necessary for study 3 were obtained by

means of a questionnaire posted on the website. The questionnaire was posted on website A between September and October 2004. It was linked through session-IDs to the corresponding logfiles of the same period.

At that time the website also offered information in French, with an interface that was identical to all other language versions. The questionnaire was embedded into the disease-information section of the website and was available in each of the website's language versions (German, English, French, Spanish, Portuguese). Users were free to decide whether to answer the questionnaire or not.

2.3.1.3 Procedure

Cultural indices (Hall, 1959; Hofstede, 1991) were assigned based on the user's origin using Geoselect software (see section 1.3.2.1.1.2). Due to divergences in the availability of cultural quantitative data for the various cultural dimensions, the amount of data used differs between the studies. Final analyses are based on the following data:

Study 1: data from 5,136 sessions and 54,074 page requests in total (55 countries).

Study 2 and 3a: data from 282,713 sessions and 2,200,317 page requests (188 countries)

Assignment to either the native or non-native speaker group was also based on geographical information, mapped here to our language database (see section 1.3.2.1.1.2). Additionally, for the L1 user group we controlled the correspondence of the user's native language and the language of the website visited. However, cases where native speakers surfed on the website in a language other than their mother tongue were rare. (We classified these cases as L2 user situations.)

In the questionnaire, a 5-point anchored Likert-scale asked non-native speakers for their proficiency level of the language in which the website was read. The same Likert-scale was also employed to assess the user's web experience. Users were asked to categorize themselves as either patients or physicians by choosing from a drop-down menu the category (*"Patients"*, *"Physicians"*, *"Other"*) that describes them best.

In all our studies we assume that the groups differed only in their linguistic or cultural characteristics, i.e. that other factors that may influence search behaviour were equally distributed within the two groups. We also assumed that a person's search behaviour was not affected by prior visits to the site, or if this was the case, first-time users and repeat visitors were equally distributed

in the two groups. These restrictions are unavoidable given anonymous data without persistent cookies. Like many other studies in website usage research, our investigation relies on large sample sizes.

2.3.2 Study 1: The Impact of Culture on Preferences for Navigational Structures

2.3.2.1 Conceptual Framework and Hypotheses

Navigation behaviour encompasses numerous different behavioural aspects. Describing, categorizing, or measuring it is therefore a complex endeavour. For example, scientists' countless efforts to describe *"lostness"* illustrates the difficulty of finding patterns for individual navigation behaviour.

For the purpose of our research we focussed on aspects of navigation behaviour that we expected to be strongly influenced by the users' cultural backgrounds. These aspects are the linearity of navigation patterns, the amount of information accessed by the user, and time spent on the website.

Cultural patterns that describe one's time and information related preferences are Hofstede's cultural dimensions of Long-Term-Orientation and Uncertainty Avoidance, as well as Hall's dimension of Poly- and Monochronicity (for a brief description of these dimensions see Appendix A-1.1).

2.3.2.1.1 Characteristics of Navigation Patterns with Regard to Culture

Information Need and Uncertainty Avoidance.

Whereas members of low Uncertainty Avoidance cultures prefer complexity with maximal content and choices, members of high Uncertainty Avoidance cultures expect structures that allow a maximum amount of predictability (Marcus and West Gould, 2000). The reason for this preference is that members of high UA cultures tend to explore all available options in order to minimize the number of unknown situations and locations. This led us to the following hypothesis:

H1: Members of high UA cultures collect more information on a given website than members of low UA cultures.

Time and Long-term Orientation.

Based on Hofstede's definition of Long-Term-Orientation, Marcus et al. (2000) believe that long-term oriented web-designers emphasize "patience in achieving results and goals". In contrast,

short-term oriented designers express a "desire for immediate results and achievement of goals" (Marcus and West Gould, 2000).

Accordingly, when navigating on a website, short-term oriented cultures should tend to only scan (as opposed to read) pages in order to reach their navigational goal as fast as possible. Long-term oriented cultures, on the other hand, should tend to read the pages' content in more detail, which is a primarily long-term oriented view of the utility of information resources. Consequently, members of long-term oriented cultures should tend to spend more time visiting a page, or should visit fewer pages per time unit. These conclusions are also in line with the theory of Information Foraging (Pirolli and Card, 1999), which describes page accesses as a trade-off between a person's (imperfect) perception of the value and costs of accessing these pages. We derived the hypothesis:

H2: Members of short-term oriented cultures spend less time on pages visited than members of long-term oriented cultures.

Linearity and Monochronicity.

Hall's dimension of Monochronicity is also a time-related dimension that refers to the way time is structured. In its simplest form, time structuring occurs either in linear (based on fixed schedules, monochronic) or non-linear (with easily interchangeable timetables, several things at a time - polychron) form. With regard to navigation behaviour, members of monochronic cultures should exhibit linear navigation patterns, which are characterized by successive access to different page sets. On the other hand, members of polychronic cultures should navigate in a non-linear way. Non-linear navigation is represented by forward and backward moves and repeated accesses to the same page set. Therefore, we formulated that:

H3: Members of monochronic cultures are more likely to show linear navigation patterns whereas members of polychronic cultures are more likely to show non-linear navigation patterns.

2.3.2.2 Method

2.3.2.2.1 Materials and Apparatus

We obtained information about navigation behaviour by using the tool WUM (www.hypknowsys.de). WUM is a sequence miner, i.e. it identifies specific access sequences of webpages. Two innovative features made this a useful tool for our study: First, it supports the search for frequent sequences that contain wildcards (for example, all paths from webpage A to webpage B, regardless of which other pages were visited in between). However, frequent patterns

also require an interpretation of the website's content in order to evaluate the usefulness of the patterns discovered. As a second essential feature, WUM therefore contains a query language (MINT) that allows the analyst to conduct a search that is based on content-based groups of pages (for example, from a diagnosis page via an index page to a second, and different, diagnosis page). Consequently it allows the analyst to narrow the search, revealing interesting patterns. Page classification is done prior to the sequence analysis. Appropriate grouping of structural and content characteristics of webpages depends on the analyst's viewpoint. For the purposes of our analysis, the pages of website A were classified as belonging to one of the three index levels or as belonging to the page set of a specific diagnosis for each topic group and language. (For instance, all pages containing information only about disease z would appear as individual URLs before page classification and as the abstracted page concept of disease z after page classification[33]).

Additional data (such as number of page requests, visit time per page) were obtained from database queries to which the logfile data had been transmitted.

2.3.2.2.2 Design: Measures

Linear and Non-Linear Navigation Sequences

Measures of navigation behaviour were based on the frequency of selected navigation sequences within the logfile partition belonging to a given country. They are defined as sequences of page accesses (e.g., Shahabi, et al., 1997) that obey *structural* as well as page-*content* constraints. The sequences are expected to differ between countries and exhibit characteristic patterns.

> *„A Web site […] can be modeled as a graph[34] […]. Navigation traverses this graph, visiting some nodes but not others, revisiting some nodes, but considering other nodes only once each in long linear sequences. In this traversal, patterns may exist." (Berendt and Brenstein, 2001)*

Unlike many other graphic analyses, in our study we use a confirmatory approach for investigating navigation patterns. Linear navigation graphs are characterized as nodes having at most one parent and one child node each (Van Dyke Parunak, 1989). Expressed differently, a linear sequence does not contain a request for a page previously visited in the same sequence. With our website, we define a navigation sequence as linear when different diagnoses are visited successively, without returning to a previously visited *diagnosis*. Yet it should be noted that linearity is not only classified

[33] For confidentiality reasons, we are not able provide examples of the taxonomy and queries applied for website A. For an abstract example of a WUM query see Appendix A-2.1

[34] A graph is composed of nodes (here: webpage) and links between them (here: hyperlinks). Conklin, E.J. (1987): Hypertext: An Introduction and Survey, IEEE Computer 20 [9], pp. 17-41.

on the structural level, but also on the content level. Since index pages primarily serve to organize and help access information pages (Neerincx, et al., 2001), we also consider navigation patterns as linear if index pages are visited between each diagnosis visit. Thus, the notion of linearity not only prohibits cycles along a user's path, but also emphasizes sequential information access. Figure 9 depicts two examples of linear navigation sequences.

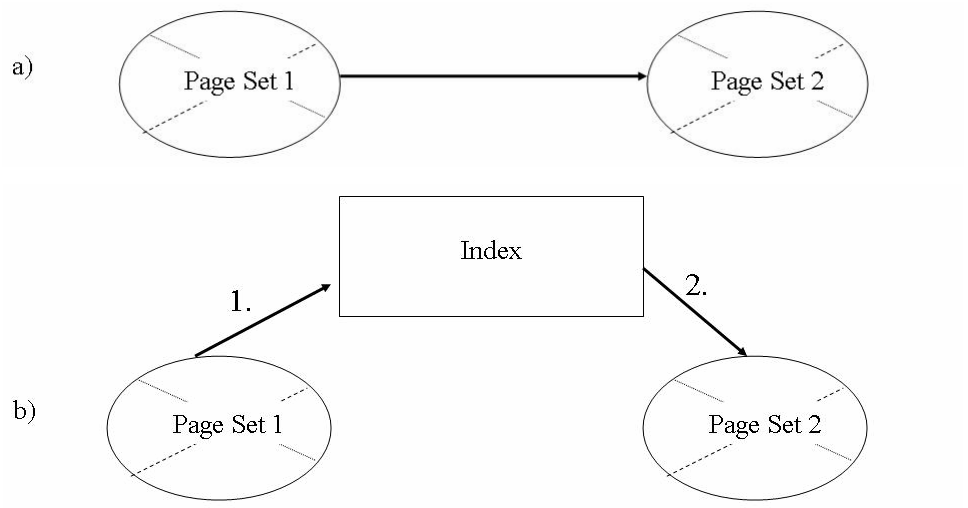


Figure 9. Examples of Linear Navigation Sequence

We classify a navigation sequence as non-linear when a diagnosis page set is visited more than once within a navigation sequence, and a navigational ring is created (see Figure 10). Similar characterizations of linear vs. non-linear local patterns have been described in a number of studies since Canter (Canter, et al., 1985). In contrast to these earlier studies, here again we do not define patterns in a purely structural way, but use abstracting concept hierarchies in order to assess the degree of "content-based" linearity.

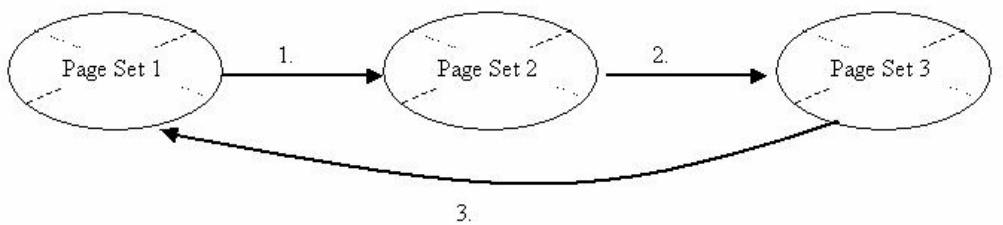


Figure 10. Example of Non-linear Navigation Sequence

(For other measures that compare navigation behaviour recorded in different logfile subsets, see for example Berendt, et al., 2002; Cadez, et al., 2000; Spilioupoulou and Pohle, 2001).

The non-linearity (respectively linearity) of different countries' users' navigation behaviour is determined as the difference between the number of non-linear and linear paths divided by the sum of both.

Other Measures

Index scores of Long-Term-Orientation and Uncertainty Avoidance for different countries were taken from Hofstede (1980; Hofstede, 1991). The classification of countries as either monochronic or polychronic follows the demographic scale suggested by Morden (1999). In order to test our hypotheses, the index scores of cultural dimensions were correlated with several measures derived from the logfile analysis, as described in the following paragraph.

We used the average time spent per page request in seconds for each country in order to test H2, referring to the relationship between Short- vs. Long-term Orientation and time spent on visited pages. The amount of information that H1 expects to be correlated with the degree of Uncertainty Avoidance was measured by the number of page sets requested, which were measured by the average number of distinct diagnoses accessed during a session.

Since countries were ordered from low values for monochronic to high values for polychronic cultures, we expect a positive correlation between this scale and the non-linearity of navigation behaviour. Monochronic countries should yield strong negative scores and polychronic cultures should yield positive scores for the difference measurements, whereas countries in the middle of the monochronic vs. polychronic scale should yield scores near zero.

Additionally, in order to consider the varying number of page requests or sessions per country, every country's data was weighted either by the number of sessions (for H1), the number of page requests (for H2), or the number of linear and non-linear paths (H3) available for each analysis.

2.3.2.3 Results

We used Pearson's product-moment correlation coefficient in order to test our hypotheses. The results are given in table 4.

Table 4. Results for the Relationships Assumed by the Hypotheses

(Pearson's correlation coefficient r)

Relationship	*r*	*Basis of data*[a]
H1: Uncertainty Avoidance (low to high) / diagnoses per session	.196***	53 countries with 5,030 sessions
H2: Short- vs. Long-term Orientation / average time spent per page request	.186***[b]	32 countries with 33,858 page requests
H3: Monochronic vs. Polychronic / difference of non-linear minus linear paths to sum of both	.190***	28 countries with 3,054 linear and non-linear paths

[a] Differences in the number of countries examined are due to differences in the number of countries for which the cultural index is provided.

[b] relationship is significant for $p < .001$

As illustrated in table 4, significant results were obtained for all three relationships under investigation. Results do yield – although not strong - consistent correlations. All results are in line with our hypotheses. As assumed in H2, members of short-term oriented cultures spent less time on pages visited than members of long-term oriented cultures. The more a country is considered to be uncertainty avoidant, the more information is collected by members of these countries, which is also consistent with our assumption. Finally, the positive correlation of the H3 test confirms that on average, monochronic cultures show more linear navigation patterns than polychronic cultures, and vice versa. We also controlled for the effects of native vs. non-native speakers, but no effect on cultural dimensions under investigation was found.

2.3.3 Study 2: The Impact of Culture on the Use of Search Options

2.3.3.1 Conceptual Framework and Hypotheses

Search behaviour involves the use of search engines as well as the way users browse through hypertext in forms of classification schemes. Analyses of search options therefore share a number of commonalities with studies of navigation patterns. We consequently chose similar cultural dimensions for this analysis. These are the users' need for information (as expressed by Hofstede's dimension of Uncertainty Avoidance and Hall's context specificity) and the value users assign to time (as expressed by their Long-term Orientation). In addition, we also selected the dimension of Power Distance as an indicator of individual's relationship with space.

2.3.3.1.1 Characteristics of Search Options with Regard to Culture

Information Need and Uncertainty Avoidance and Context Specificity

In line with previous studies, we propose that divergences in the amount of information needed are due to differences concerning Uncertainty Avoidance but also context specificity. We expect high UA and context specificity to be similar in their effects, both leading to a more extensive collection of information (see also results from study 1). Since navigation through hypertext provides more context information than search engine results do, we suggest that members of low-context cultures and low-UA cultures tend to opt more for the use of search engines than members of high-context and high-UA cultures do. Browsing through a collection of information, as opposed to using a search engine or alphabetical index, is an activity that is characterized by the inclusion of many resources. It therefore allows an evaluation of the various items found, thus reducing the ignorance or uncertainty one has about a topic area (Rice, et al., 2001, p. 302). Since content-organized links provide the most context information among all search options, we suggest that this type of hyperlinks is preferred by members of high-context cultures and by members of high-UA cultures. A similar suggestion was made by Luna et al. (2002), who suggested, in a theoretical framework, that a hierarchical structure is preferred by high-context cultures and search-based structures by low-context cultures.

These propositions led us to the following four hypotheses:

H4: Members of low-context cultures have a stronger preference for search engines than members of high-context cultures.

H5: Members of low-UA cultures have a stronger preference for search engines than members of high-UA cultures.

H6: Members of high-context cultures have a stronger preference for content-organized links than members of low-context cultures.

H7: Members of high-UA cultures have a stronger preference for content-organized links than members of low-UA cultures.

Time and Long-term Orientation

With regard to navigation behaviour, we took Hofstede's and Hall's time-related cultural dimensions into account, and only considered the aspect of Long-term Orientation.

When searching for information, short-term oriented cultures are assumed to opt more often for the use of the search engine than long-term oriented cultures, due to the value they assign to time. Long-term oriented cultures, on the other hand, should exhibit stronger preferences for hyperlink navigation, which require more patience to achieve navigational and functional goals, but also enables the browsing user to evaluate whether items will be of interest (Rice, et al., 2001, p.256) – a more long-term oriented view of the utility of information resources than a search for a specific, known item.

In a similar manner, we believe alphabetically organized links to be preferred by short-term oriented cultures than by long-term oriented cultures. This is due to the fact that alphabetically organized links are perceived as offering a more direct and hence faster access to the desired information than content-organized hyperlinks.

According to these statements we hypothesize that:

H8: Members of short-term oriented cultures have a stronger preference for search engines than members of long-term oriented cultures.

H9: Members of cultures with Long-term Orientation have a stronger preference for content-organized links than members of cultures with Short-term Orientation.

Space Perception and Power Distance

Within the context of cultural dimensions, the issue of Power Distance is the aspect that appears to have the most important impact on space perception. As indicated by Marcus, members of high-PD cultures are characterized by taller hierarchies in their mental models, and by a greater ability and desire to structure their access to information (Marcus and West Gould, 2000). In the context of information seeking behaviour, users from low-PD cultures should consequently tend to avoid following links through a deep hyperstructure and therefore favour the use of search engines (and vice versa for members of high-PD cultures).

Thus, we formulate the following hypothesis:

H10: Members of high-PD cultures have a stronger preference for hyperlink searches than members of low-PD cultures.

2.3.3.2 Method

2.3.3.2.1 Design: Measures

The preference exhibited for a search option was measured by three variables: (a) the percentage of users within one group who used the search option, (b) how often the search option was used within one session (absolute frequency), and (c) the frequency of use relative to the number of page requests. Higher preference is indicated by a higher percentage of users and higher absolute and relative frequency of use. Preference is generally an attitudinal construct that usually results, however, in behavioural consequences. The selection of these measures is based on an assumed two-step decision process. The first measure tests whether a particular search option was used at all (first step). If one search option is not used at all it usually indicates either strong indifference towards this particular search option or a strong preference for another search option. If a search option has been used once, the user decides whether or not to continue using this tool (second step). The frequency of its use is hence an indicator of its ease of use and information access barriers. For this reason the measure of absolute frequency was chosen. Finally, the third measure normalizes results from the second measure by the number of accessed pages. The analyst needs to decide whether absolute or relative frequency is more important. Due to difficulties assessing users' situational backgrounds in such a large quantitative study, both variables have their own justifications. The term "preference" is used here as a collective term describing all of the measures employed.

2.3.3.3 Results

Between-(Cultural Groups) Comparison:

Table 5 shows the selection of search options by cultural dimension. (Exact data can be found in Appendix A-2.2.) Concerning three cultural dimensions out of four, evidence from our data indicated that the users' selection of search options occurs in the predicted way. Largely, the first two measures (search options used or not; number of uses) confirm our hypotheses, whereas the third measure reveals opposite results. Uncertainty Avoidance is the only cultural dimension with inconclusive results that largely contradict our hypotheses.

Table 5. Selection of Search Options Ordered by Cultural Groups

	UA	*Context*	*LTO*	*PD*
Search Engine				
Measure 1	HIGH	LOW	LOW	
Measure 2	HIGH	LOW	n.s.	
Measure 3	LOW	LOW	n.s.	
Alphabetical Search				
Measure 1				**HIGH**
Measure 2				**HIGH**
Measure 3				LOW
Content Search				
Measure 1	LOW	HIGH (p=0.075)	**HIGH**	**HIGH**
Measure 2	n.s.	n.s.	**HIGH**	**HIGH**
Measure 3	n.s.	HIGH	LOW	**HIGH**

Groups with higher preference are displayed in the cells.

Measure 1: Percentage of users who used the search option;

Measure 2: absolute frequency of use;

Measure 3: relative frequency of use (Only users who used at least one search option were considered.)

Grey /hatched: no hypothesis about the distribution

Bold: confirms our hypotheses

n.s. = differences are not significant

Uncertainty Avoidance (H5): Contrary to our hypothesis, a significantly higher percentage of users within the high Uncertainty Avoidance group use the search engine compared to the low Uncertainty Avoidance group. Results concerning the use of a content search are significant but contradictory.

Context Specificity (H4): In line with our hypothesis, low context users prefer the use of search engines more than high context specific users, whereas it is the opposite for the use of content searches. Results regarding the use of the search engine are significant, those regarding the use of the content search exhibit a tendency in the right direction, yet without significance.

Long-Term-Orientation (H8 & H9): The first two measures reveal a higher preference for search engines among users from short-term oriented cultures, and a higher preference for content search among users from long-term oriented cultures. These outcomes are in line with our hypotheses.

Power Distance (H10): Results reveal significant differences between low and higher power distant users with regard to their use of hierarchical searches (i.e. alphabetical search and content search). Confirming our hypothesis, hierarchical searches are preferred by high power distant users.

Further General Findings. Although we did not test the results for within-group differences, the numbers show a consistent order in search option preferences: across all groups, alphabetically organized links were more popular than content-organized links. These in turn were much more popular than the search engine.

2.3.4 Study 3: The impact of language on the use of search options and the role of domain knowledge as mediating factor

2.3.4.1 Conceptual Framework and Hypotheses

In contrast to previous cultural studies that are founded on thinking patterns, our analysis of language is based on users' cognitive effort[35]. In line with Navarro-Prieto and Scaife (1999) and with the Information Foraging Theory (Pirolli and Card, 1999), we suggest that dissimilarities in information access are due to differences in cognitive effort and terminological knowledge (semantic network) as presented in section 1.2.2.1.1.5. Differences exist between native speakers and non-native speakers, between receptive and productive language use, and between speakers of diverging levels of topic knowledge.

Accordingly, in this study we compare native speakers and non-native speakers of the languages in which a website's information is presented, as well as users of low and high domain knowledge within each group.

2.3.4.1.1 Characteristics of Search Options with Regard to Language

Like our previous study, this investigation of language-determined search behaviour is based on the assumption that the characteristics of various search options correspond to users' linguistic abilities to different degrees. In the introductory language section we elaborated on reasons for differences in cognitive effort: These were the acquisition of language as a native or non-native language (RHM), the use of receptive or productive language, and the differentiation of (specific parts) of the semantic network – also dependent on the speaker's domain knowledge.

---

[35] There is no clear line between "thinking patterns" and "cognitive effort". The term "thinking patterns" emphasizes a habit, but are also closely linked to cognitive effort. Thinking patterns express an habitual and therefore automised way of thinking. Automised ways of thinking require less cognitive resources and therefore less cognitive effort.

We expect that due to its effects on search behaviour, domain knowledge can mediate the effects of lower language proficiency.

Our investigation of search behaviour is based on the assumption that the characteristics of various search options correspond to different degrees to users' linguistic abilities. Typical traits of three common types of search options – search engine, alphabetical search, and content-organized hyperlinks – are therefore examined.

Search Engines

Initiating a search with a search engine requires active language capacity: In order to successfully retrieve information from a search engine the user has to enter one or more keywords that best match the information sought. These words all have to be generated from the user's productive vocabulary knowledge.

Furthermore, the ability to formulate appropriate queries is closely related to the density of the semantic network of the user's active vocabulary. A user with insufficient language ability has to employ more general search terms, hardly being able to add synonyms and parallel terms in order to refine the search field. If the user's goal is to maximize precision rather than recall (i.e., if she/he prefers to find pages that are *particularly* relevant to the query rather than finding *all* pages that may be relevant), the semantic imprecision of the search word used may lead to few or zero matches, a disappointing outcome.

Alphabetical Search.

From a linguistic point of view, the use of an alphabetical search hardly differs from the use of a search engine. Like search engines, alphabetical searches often require productive vocabulary knowledge. In most cases the number of alphabetical links is too high to allow the user to obtain an exhaustive overview of available terms. Once again the user has to generate a term in his/her mind and check the availability of the term among existing hyperlinks. If the term is not available, the user has to generate synonyms. We argue therefore that similarly to the use of search engine, a search by means of alphabetical hyperlinks requires productive language use and a broad terminological knowledge.

The requirements of productive vocabulary can be lower than with a search engine because starting from a word whose first letters one remembers correctly can lead to the correctly spelled word by browsing through the words before and after the first guess. In addition, selecting a hyperlinked

word in a list requires less effort than typing it. (However, these browsing advantages disappear when one remembers a part of the word other than its beginning, or when lists become very long.) When faced with contents in a language with diacritic marks, selecting obviates the need to search for special characters that are only available in software but not on the keyboard. It may therefore be argued that an alphabetical search can rely to a higher extent on receptive language use than search engines, and an alphabetical search may be regarded as a mixed search option.

Content-organized Hyperlinks

In contrast, using hyperlinks that are organized by content allow the user to rely mostly on passive language knowledge. Content-organized links are usually arranged in a way that the user can overview all of the links available within a certain category. Therefore, the user does not need to productively recall words: the available hyperlinks are helpful cues that turn information processing into a passive and thus easier use of language.

In addition, the context information provided makes information processing easier when the user's semantic network consists of fewer nodes. In particular, links that are integrated into text or that are presented with clear visual cues provide a large amount of contextual information that facilitates language understanding. Clickable maps provide a helpful organization of contents by context: Users do not need to name the content or specify the context; they can search for the contents within its domain context and recognize both of them visually. Once they have made a selection, they are faced with either one or few more detailed content choices that can generally be differentiated with little terminological knowledge.

2.3.4.1.2 Hypotheses

According to the Theory of Information Foraging we suggest that the use of search options is determined by the costs of using them. As argued previously, the use of different search options requires different linguistic abilities and cognitive effort for native and non-native speakers. Following efficiency rules exclusively, both groups would select content-organized links since this search option involves the least cognitive effort for them.

These considerations lead us to expect that:

H11: Non-native speakers exhibit a stronger preference for content-organized hyperlinks than for search engines and alphabetically organized hyperlinks.

H12: Native speakers exhibit a stronger preference for search engines and alphabetically organized hyperlinks than non-native speakers do.

Figure 11 shows how these hypotheses circumscribe the area of search behaviour that we expect to be sensitive to language proficiency. The area shown in dark grey is characterized by higher cognitive effort, and therefore higher costs of information processing. The arrows emphasize that H13 is about expected differences within the L2 user group, while H14 concerns expected differences between the L2 and the L1 groups.

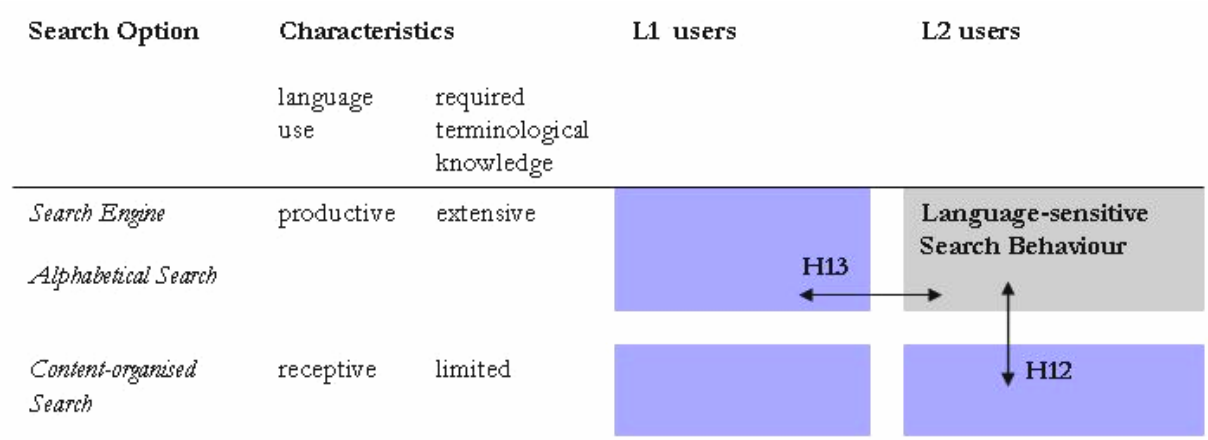


Figure 11. Overview of Hypothesized Language Sensitive Search Behaviour

We expect that domain knowledge mediates language effects so that non-native speakers with high domain knowledge will incur lower access costs than non-native speakers with low domain knowledge, due to differences in language proficiency. Since differences in proficiency levels are more salient among non-native speakers, the impact of domain knowledge is more important in the L2 user group. We therefore expect the following outcome:

H13: The percentage of users with high domain knowledge is higher among non-native speakers than among native speakers.

H14: Non-native speakers with high domain knowledge exhibit a stronger preference for search engines and alphabetical searches than non-native speakers with low domain knowledge.

2.3.4.2 Study 3a: A Logfile-based Analysis of Language and Search Behaviour

2.3.4.2.1 Method

Design:: Measures

The linguistic diversity of the users and their large number allowed us to compare L1 and L2 users, with each group representing a large variety of linguistic backgrounds.

We determine the preference for the various search options based on the same measures as in the study 2. Preferences for search options are analysed between the two language groups (L1 vs. L2) as well as within the L2 group, depending on the hypothesis.

Website access was determined by the number of users who requested at least one page on the website. Logfile data were inserted into a relational database (MySQL). All data were then analysed using SPSS.

2.3.4.2.2 Results

Search option preferences: native speakers vs. non-native speakers (hypotheses 11 & 12)

The data shows that a higher percentage of L2 users than of L1 users employed at least one search option (38.7% vs. 30.6%). The difference was significant ($p < 0.001$ in a χ^{2}-test).

Table 6 gives an overview of the results obtained from univariate MANOVAs. The table should be read as follows: ** indicates that the difference was significant at the 1% level, * that it was significant at the 5% level.

Table 6. Exhibited Search Preferences by Language Group in Study 3a

	L1	*L2*
Search engine		
Measure 1	29.3%	**30.8%****
Measure 2	1.77 (1.868)a)	**1.82**** (1.900)
Measure 3	0.17 (0.186)	**0.20**** (0.237)
Alphabetical search		
Measure 1	56.0%	**57.6%****
Measure 2	3.80 (4.983)	**4.21**** (5.121)
Measure 3	0.31 (.283)	0.25 (0.198)

Content search		
Measure 1	**47.6%**	45.3%**
Measure 2	1.52 (1.246)	**1.65*** (1.738)
Measure 3	0.19 (0.235)	0.15 (0.164)

Measure 1: %age of users who used the search option;

Measure 2: absolute frequency of use;

Measure 3: relative frequency of use (Only users who used at least one search option were considered.)

Bold: Language group with significant higher preference.

** significant at the 0.001 level; * significant at the 0.05 level

The data from the study indicates search option use that either contradicts our expectations or is inconsistent across the different measures. The results for the first measure (search option used or not) were $F(1,277802) = 359.15$, $p < 0.01$ (search engine), $F(1,277802) = 629.22$, $p < 0.01$ (alphabetical search), and $F(1,277802) = 239.02$, $p < 0.01$ (content search). For the absolute frequency of use, we obtained: $F(1,88303) = 13.50$, $p < 0.01$ (search engine), $F(1,88303) = 63.81$, $p < 0.01$ (alphabetical search), and $F(1,88303) = 3.976$, $p < 0.05$ (content search). The results for the relative frequency of use were $F(1,12148) = 12.07$, $p < 0.01$ (search engine); $F(1,12148) = 0.80$, $p = 0.373$ (alphabetical search), and $F(1,12148) = 1.63$, $p = 0.202$ (content search).

Within-L2-group comparison (H11):

Within the L2 user group, none of the three measures indicated a higher preference for content search compared to the other two search options. A higher percentage of L2 users opted for the alphabetical search. Furthermore, within this group the content search was the least-used search option, both absolutely and relative to the number of page requests.

Hypothesis 11, predicting a preference for content search over alphabetical searches and search engines among the non-native speakers, could therefore not be confirmed.

Between-group comparison (H12):

Most measures suggest that L2 users have a stronger preference for search engines and alphabetical search than L1 users do (the difference in relative frequency of alphabetical searches is in the expected direction, but it is not significant). The results for preferences for content-organized searches are inconsistent with one another.

Hypotheses 12 could therefore not be confirmed.

Further general findings:

Alphabetical searches were always used most often. With regard to the absolute frequency of use (how often the search option was used within one session), all outcomes showed a higher usage of all search options by non-native speakers. This result is in line with the higher number of page requests and the higher percentage of search option users among non-native speakers.

Following arguments and findings from previous research, the percentage of physicians can be expected to be higher among the non-native speakers than among the native speakers.

Results from a survey that was conducted on the website between October 2000 and October 2002 revealed a lower percentage of physicians among the users from North and Middle America (45,1%), Europe (59,3%)[36] and Australia (58,8%) than among the users from Africa (74,1%), Asia (79,7%) and South America (80,8%). With the exception of South America, the parallel distribution of physicians and L2 users confirms our expectations.

Based on the results of study 3a alone, this interpretation is only conjecture since the Web server logs do not allow us to measure the personal variable "domain knowledge".

The use of logfiles also involved a number of methodological challenges as described in section 1.3.2.1.1

2.3.4.3 Study 3b: A Logfile-Plus-Questionnaire Based Analysis of Language, Domain Knowledge, and Search Behaviour

2.3.4.3.1 Materials and Apparatus (Additional)

In order to test the role of domain knowledge in language-related use of search options, we conducted a follow-up study where in addition to using logfile data, we assessed the user's medical knowledge by means of a questionnaire. Users were asked to rate themselves as *"patients*", *"physicans"*, or *"others"*.

Due to the introduction of session IDs and the resulting possible linkage between questionnaires and logfile data, we were able to skip the major part of logfile data preparation and cleaning. Unless

[36] The majority of the European users (68%) came from L1 countries and cities.

a user answered the survey twice with diverging data, one session corresponds to exactly one unique user. Each session is again characterized through the pages visited and the language group.

2.3.4.3.2 Design

As a second independent variable, the user's domain knowledge is introduced. *"Patients"* were treated as users with low domain knowledge, and physicians/medical personnel as users with high domain knowledge.

All dependent measures of search behaviour were the same as in previous studies.

2.3.4.3.3 Results

Data preparation and basic statistics. 30 sessions with questionnaire answers had to be excluded due to data problems or impossible assignment to either the patient or physician group, leaving 135 valid cases.

40,2% of the participants were male, 59,8% female. The average age was 39.49 years (Std.dev.: 11,68) with a minimum age of 18 and a maximum age of 67. Internet experience covered all offered five categories (1 = *"not more than once a week"* – 5 = *"almost always online"*) and averaged out at 2,77 (Std.dev.: 1,17), which corresponds to *"online several times per week"* (=2) or *"daily for a few minutes"* (=3).

No significant differences were found between the web experience ($p = 0.105$, Mann-Whitney's U-test), age ($p = 0.169$), or gender ($p = 0.177$) of native and non-native speakers. Standard weighting procedures were applied to reflect the different numbers of native and non-native speakers. The two groups of different domain knowledge do not differ significantly with regard to age ($p = 0.602$) or Internet experience ($p = 0.147$). The percentage of women is significantly higher in the low domain knowledge group ($p = 0.009$).

In study 3b, native language was derived from the users' self-reports. Cross-validation showed that the results of software-based native-language classification differed from self-reports in 6.7% of the cases.

On the 5-point-Likert-scale, non-native speakers rated their language knowledge with a mean of 4.18 (standard deviation: 1.14), which corresponds to a rather high proficiency level (Fluent speaking and understanding).

96 of the participants were patients (71.1%) and 13 were physicians (9.6 %). 26 people could not be assigned to either of these groups (see above).

The 26 participants who classified themselves as "other" with respect to medical domain knowledge were excluded from analyses pertaining to domain knowledge.

Search option preferences: native speakers vs. non-native speakers (hypotheses 11 & 12). The data show that a higher percentage of L2 users than of L1 users employed at least one search option. The difference was even more pronounced than in study 3a (68.2% vs. 40.7%). The difference was significant ($p < 0.05$ in a χ^2 -test).

Within-L2-group comparison (H11) and between-language-group comparison (H12):

Table 7 gives an overview of the results obtained from MANOVA.

Table 7. Exhibited Search Preferences by Language Group in Study 3b

	L1	*L2*
Search engine		
Measure 1	44.4%	**83.3%****
Measure 2	0.54 (1.605)	**2.63**** (6.776)
Measure 3	0.06 (0.155)	0.09 (0.144) (p = 0.392)
Alphabetical search		
Measure 1	77.1%	66.7% (p = 0.168)
Measure 2	1.94 (5.171)	**4.21*** (7.892)
Measure 3	0.06 (0.155)	0.09 (0.144) (p = 0.392)
Content search		
Measure 1	44.4%	50.0% (p = 0.286)
Measure 2	1.94 (5.171)	5.53 (18.234) (p = 0.069)
Measure 3	0.07 (0.237)	0.06 (0.196) (p = 0.963)

Numbers in ( ) show standard deviation.

Measure 1: %age of users who used the search option;

Measure 2: absolute frequency of use;

Measure 3: relative frequency of use (Only users who used at least one search option were considered.)

Bold: Language group with significant higher preference.

** significant at the 0.001 level; * significant at the 0.05 level

Similar to study 3a, the data from study 3b does not provide evidence of clear preferences for one of the search options. Results are mainly inconclusive given non-significance or inconsistent differences. For the first measure, $F(1,118) = 6.60$, for the second, $F(1,118) = 4.38$ and for the third, $F(1,118) = 3.35$. Again, neither hypothesis 12 (between-language group comparison) nor 11 (within-language-group comparison) could be confirmed.

Search option preferences: the influence of domain knowledge (hypotheses 13 and 14). The majority of the survey participants were patients. However, the percentage of physicians differed significantly between L1 and L2. Among the 93 participants that were patients, 83 were native speakers (= 87.4%), and 12 (=12.6%) were non-native speakers. The percentage of physicians among the L2 group (25.0%) was significantly higher than among the L1 group (7.8%) (χ^2 -test, $p < 0.05$), which confirms hypothesis 13.

The following table shows means and standard deviations regarding differences between physicians and patients, grouped by their language proficiency.

Table 8. Exhibited Search Preferences by Language Group and Domain Knowledge in Study 3b

	L1		*L2*	
	Physicians	*Patients*	*Physicians*	*Patients*
	N = 7	*N = 83*	*N = 4*	*N = 12*
% of users that used at least one search option:				
	28.6%	**43.4%**	**100%**	58.3%
Search engine				
Measure 1	**50.0%**	44.4%	**100%**	71.4%
Measure 2	0.14 (0.378)	**0.58** (1.668)	**9.25** (13.841)	0.67 (.985)
Measure 3	0 %	**44.4%**	**75%**	42.9%
Alphabetical search				
Measure 1	50.0%	**77.8%**	**75%**	57.1%
Measure 2	0.29 (0.756)	**2.41** (5.871)	**12.75** (13.276)	2.08 (4.379)
Measure 3	0.007 (0.018)	**0.076** (0.175)	**0.23** (0.221)	0.05 (0.102)
Content search				
Measure 1	0 %	**44.4%**	75%	42.9%

Measure 2	0 (0.000)	**2.44** (6.919)	3.75 (2.986)	7.50 (22.976)
Measure 3	0 (0.000)	**0.089** (0.271)	0.08 (0.066)	0.08 (0.245)

Numbers in ( ) show standard deviation.

Measure 1: percentage of users who used the search option;

Measure 2: absolute frequency of use;

Measure 3: relative frequency of use (Only users who used at least one search option were considered.)

Bold: **Domain knowledge group** (within a language group) with significant higher preference.

Results obtained from two-way MANOVA (Pillai's Trace), regarding the impact of domain knowledge and language, reveal the following. "SE" is used as an abbreviation for "search engine", "AS" for alphabetical search, and "CO" for content-organized-search.

The data shows, for the absolute (Measure 2) and relative frequencies (Measure 3), significant effects of domain knowledge (SE absolute: $F(1,117) = 17.22$, $p<0.01$; SE relative: $F(1,118) = 6.90$, $p = 0.01$; AS absolute: $F(1,117) = 6.13$, $p < 0.05$), a significant effect of language (SE absolute: $F(1,117) = 21.94$, $p<0.01$; SE relative: $F(1,118) = 13.87$, $p = 0.01$; AS absolute: $F(1,117) = 12.39$, $p < 0.01$; AS relative: $F(1,118) = 4.93$, $p < 0.05$), and significant interaction effects (SE absolute: $F(1,117) = 21.17$, $p < 0.01$; SE relative: $F(1,118) = 10.30$, $p < 0.01$; AS absolute: $F(1,117) = 13.89$, $p < 0.01$; AS relative: $F(1,118) = 7.56$, $p < 0.01$). The only exception is the non-significant impact of domain knowledge on the relative use of alphabetical search (AS relative: $F(1,118) = 1.48$, $p = 0.226$).

No significant impact was found for content search (domain knowledge – CO absolute: $F(1,117) = 0.896$, $p = 0.346$; domain knowledge – CO relative: $F(1,118) = 0.346$, $p = 0.495$; language – CO absolute: $F(1,117) = 1.816$, $p = 0.180$; language – CO relative: $F(1,118) = 0.248$, $p = 0.640$; interaction – CO absolute: $F(1,117) = 0.040$, $p = 0.841$; interaction – CO relative: $F(1,118) = 0.7559$, $p = 0.570$.

With respect to the percentage of users who used a certain search option, we obtained the following data. Domain knowledge significantly affects only whether or not a search engine has been used ($F(1, 118) = 5.325$, $p < 0.05$), whereas the effect of language (SE: $F(1, 118) = 21.872$; $p<0.01$; AS: $F(1,118) = 5.086$, $p = 0.026$; CO: $F(1,118) = 12.641$, $p<0.01$) and the interaction (SE: $F(1, 118) = 7.505$, $p=0.007$; AS: $F(1, 118) = 5.222$, $p = 0.024$; CO: $F(1, 118) = 9.310$, $p = 0.003$) were significant for all search options.

The influence of domain knowledge in the L2 language group: Within the L2 group, all measures show that physicians preferred the use of a search engine and alphabetical search more

than patients. In line with the results above, one-way MANOVA revealed significant differences pertaining to the use of search engines and alphabetical search (SE: used/non used: $F(1,30)=10.500$, $p<0.05$; absolute use: $F(1,30)=11.326$, $p<0.05$; $F(1,30)=7.88$, $p<0.05$; AS: used/non used: $F(1,30)=4.573$, $p<0.05$; absolute use: $F(1,30)=13.844$, $p<0.001$; relative use: $F(1,30)=11.249$, $p<0.05$). In contrast, differences pertaining to the use of content search were not significant. Hypothesis 14 can therefore be confirmed.

The influence of domain knowledge in the L1 language group: In the L1 user group differences in search option preferences between physicians and patients were not as clear. All search options appear to be preferred by the patient group.

It should be noted, however, that due to the imbalanced distribution of user characteristics across the data sets, analyses are based in some groups on a small number of participants. Low F-values pertaining to the use of content searches are an indicator.

The questionnaire in which the follow-up study assessed the users' domain knowledge increased the validity of the logfile analyses.

2.3.5 Study 4: The Impact of Language (and Culture) on Preferences for Information Categorization

Whereas the previous studies focussed on behavioural issues, this last study aims to link these findings to attitudinal aspects of language. It also links, to a certain degree, linguistic and cultural issues. The study differs furthermore from the preceding examinations by featuring an exploratory experimental approach for parts of the analysis. For the purpose of this study the effect of language was investigated from two key perspectives.

2.3.5.1 Conceptual Framework and Hypotheses

2.3.5.1.1 Culture and Language and their Impact on Information Categorization.

Within this study we tested the extent to which a mother tongue influences the users' preferences for search cues and search tools. Expected differences in information categorization are based on the following theoretical background, which were discussed in detail earlier in this thesis:

In the introductory section, links between the environment and the organization of a language's lexico-semantic structure as well as between a language's structure and a native speaker's thinking patterns were established. As such, they determine individuals' awareness for certain things and their preferences for structuring the social and physical world (see also section 1.2.2.1.1.5).

In line with results from our previous study, preferences in search cues are furthermore expected to be affected by the user's language skills. Divergences between native speakers and non-native speakers are based on differences in cognitive effort, time consumption, and vocabulary knowledge. The characteristics of search options and search cues lead to a trade-off between precision and time savings on the one hand and medical and verbal knowledge on the other hand (see section 2.3.4).

According to results from study 3, it can be expected that search options that require a higher degree of verbal skill are preferred by L1 users more than by L2 users (Kralisch and Berendt, 2004), as well as those who are the domain experts.

Research Proposition: Users of different cultural and linguistic backgrounds differ in their preferences for information categorization.

2.3.5.1.2 Effects of Differences in Information Categorization Preferences.

We use this model as an underlying concept for the comparison of test groups.

We expect the perception of a classification's usefulness to depend on the congruity of the categorization scheme with the user's culturally and linguistically determined values.

Classification schemes are a collection of categories used to describe and place a specific type of information. Yeo, for example, tested food categorization schemes that classified food either by ethnic groups or by type of food (Yeo and Loo, 2004). Hence, congruity with the categorization scheme describes how much a given classification scheme matches an individual's way of describing and placing information.

Furthermore, results from Luna & Peracchio (2002; Luna, et al., 2003) suggest that cultural congruity is a determinant of cognitive effort. An E-Health-website's categorization scheme is therefore expected to influence perceived *"ease of use"*. In a similar manner, linguistically determined cognitive effort is assumed to affect *"ease of use"*. In Davis' TAM model, *"ease of use"* is itself a determinant of *"usefulness"*.

The following hypotheses are inferred:

H16: A higher degree of congruity between a website's information categorization scheme and the user's preferences results in higher perceived ease of use.

H17: A higher (lower) degree of congruity between a website's information categorization schemes and the user's preferences results in higher (lower) perceived usefulness.

H18: A higher (lower) degree of congruity between a website's information categorization schemes and the user's preferences results in a more positive (negative) attitude towards using these search cues.

H19: A higher (lower) degree of congruity between a website's information categorization schemes and the user's preferences results in a higher (lower) use of search options.

Domain knowledge was not integrated into this study for two reasons. First, due to its effect on cultural thinking patterns and linguistic semantic networks, domain knowledge would represent another dimension in the comparison, resulting in a more complex research design. Second, due to a weaker influence of internationalized medical education (e.g. WHO ICD), non-physicians are less influenced by international medical knowledge. Their categorization preferences are therefore expected to be affected more by their cultural backgrounds, and hence to differ more among each other.

2.3.5.2 Method

2.3.5.2.1 Participants

The sample of 141 participants consisted of 32 German students, 20 Russian students, 11 British students, and 78 Malaysian students of various ethnicities (mainly Chinese, Malay, Iban). All participants spent the major part of their lives in their countries of origin. The majority of the students were between 20 and 26 years old. Those participants that were tested with a non-native language version of the website had a fair knowledge of the tested language, English for the Malaysian group and German for the Russian sample group. No significant differences existed with regard to Internet use: participants had been using the Internet for about six to seven years and 13 hours per week (British group: 6.38 hours). Participants of all groups studied non-medical subjects and rated their medical knowledge on the average as *"regular"* to *"low"*, which corresponds to their ability to name a few common dermatological diseases, sometimes including their symptoms.

2.3.5.2.2 Materials and Apparatus

Data was obtained in laboratory experiments conducted at a major German and a major Malaysian university. A number of pilot tests with each subsample were carried out prior to the experiments.

Each participant at the German university received a 9€ cinema voucher. Each participant at the Malaysian university received an entry into a draw.

2.3.5.2.3 Design

We opted for a 2x2 design. The participants of two of the four groups were native speakers of the website's languages (German and English), i.e. the treatment was linguistic correspondence (L1); the participants of the other two groups had to navigate in a language other than their native tongue (German for the Russian participants, English for the Malaysian participants). This is the L2-group. The second treatment was in the language version of the website: the experiment was conducted in two language versions (German=G and English=E), with one native and one non-native speaker group for each version.

We consequently had four groups: group 1 was an L1/G group (L1 website+German version), group 2 an L1/E group (L1 website + English version); group 3 an L2/G group (L2+German version); group 4 an L2/E group (L2+English version).

Measures are defined as follows.

Preferences. Preferences for categorization schemes were measured through analysis of criteria employed for card sorting as well as through the participants' ranking order of search cues. These search cues were: *symptoms* (e.g. swollen, rash), *colour* and *body parts* as visual search cues, *name* of the disease (e.g. acne, leprosy, neurodermatitis), and *cause* (e.g. innate, age-related, viral) as search cues that required more extended medical and linguistic skills.

By means of manual data revision we checked whether these criteria were used for classifying the card groups. Each term/description was assigned to one of the groups by one of the analysts unless a group assignment was not possible. Only a few descriptions did not match any of these categories. Participants were completely free in their choice of classification cues. They were not influenced by the search cues proposed in the questionnaire, since the ranking task followed the categorization task. The ranking task was not visible to the participants before the categorization task was accomplished.

Belief about Ease of Use, Usefulness, Attitude towards Using. Questions and measures (6-point semantic differential) are adapted from Davis (1993) (and therefore Ajzen and Fishbein (1977; 1980), and are based on his definitions . Davis defines *"attitude toward using"* as "the degree of evaluative affect that an individual associates with using the target system in his or her job".

"Perceived usefulness" is defined as "the degree to which an individual believes that using a particular system would enhance his or her [job] performance"; *"perceived ease of use"* as "the degree to which an individual believes that using a particular system would be free of physical and mental effort" (Davis, 1993), (see also Ajzen and Fishbein, 1977; Davis, 1989; Goodwin, 1987; Gould, et al., 1991; Hill, et al., 1987).

Questions about the *"ease of use"* and *"usefulness"* were asked for each search option that was used by the participants. Questions concerning their attitude towards using a certain categorization scheme regarded, beside the existing search options, search possibilities that would use other criteria such as *colour* or *non-dermatological symptoms* (e.g. fever, etc., see above). More detailed information about the questionnaire can be found in the Appendix A-2.3.

Use. Similar to the studies above, we examined how many participants within each group used a particular search option, and also controlled which search option was used first and additionally asked participants why they started with this search option.

Control Data. Furthermore, demographic data was assessed to control for age, gender, language proficiency, ethnicity, Internet use, and medical knowledge.

2.3.5.2.4 Procedure

The study combined two common data collection techniques: card sorting and a survey based on a search task. Card sorting was used to assess preferences of information categorization in an exploratory way (Dunckley and Kuldip, 2004). Card sorting techniques were shown in several studies to be efficient in detecting how people organized information (Nielsen, 1993). In usability studies in particular, this technique was used to build up meaningful structures and improve efficient navigation on a website (Lisle, et al., 1998). Results from Dunckley provided the first evidence that card sorting techniques are also useful when investigating cross-cultural differences in information categorization (Dunckley and Kuldip, 2004).

In the first step of the experiment, participants were given 30 cards with pictures of dermatological diseases, disorders, and abnormalities. The pictures were chosen in a way that the participants had the possibility to group these cards according to a number of criteria. The appropriateness of the pictures chosen was tested in pilot tests. For identification reasons a number was written on the back of the card. It was pointed out to the participants that the numbers were absolutely random. Participants were asked to sort the cards into a maximum of six groups, to name each group and to give reasons for the name choice. After card sorting, categorization preferences were also assessed

by asking the participants to rank a number of search cues according to their appropriateness for finding information about dermatological diseases.

In the following step, the test participants were asked to complete a search task that required them to navigate through website A. Group 1 (L1/G) and group 3 (L2/G) had to use the German interface whereas group 2 and group 4 (L1/E and L2/E) had to navigate through the English version.

In the last part of the experiment participants were asked, by means of a questionnaire, to rate their perceived ease of use, usefulness, attitudes, and use of the various search options. The questionnaire was based on David's work and available to the participants in English and German (For details of the questionnaire see Appendix A-2.3).

2.3.5.3 Results

2.3.5.3.1 H15: Categorization Preferences

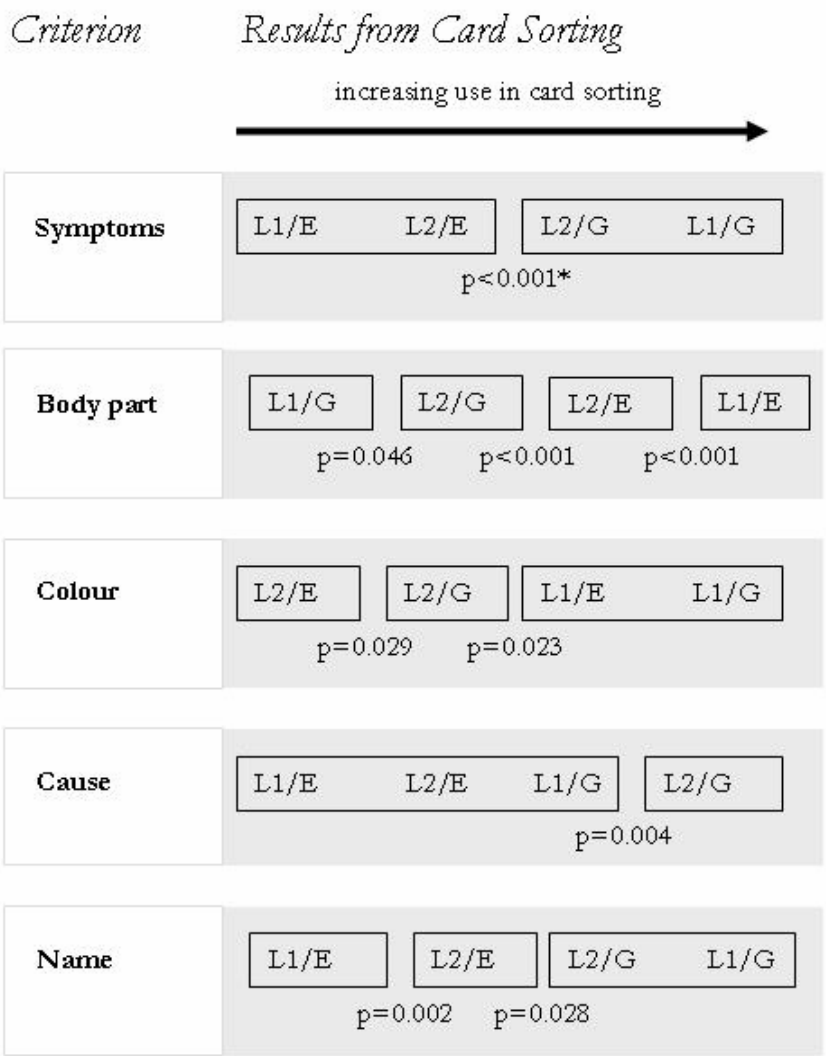
Statistical analyses. Since the goal of our analysis was to outline differences between the groups rather than provide evidence for the attitudinal and behavioural path of system use, we refrained from classical path analyses. Instead, between-group and within-group-analyses were carried out. Congruity with the groups' preferences is examined through a comparison of results regarding the various constructs. Calculations (mainly Kruskal-Wallis) are weighted by the number of participants per group.

All groups. The Kruskal-Wallis-tests of the card sorting results and the ranking order of the search cues revealed significant differences for all search criteria. Resulting from the analysis of all groups, sorting criteria can be clustered intro three groups: 1. body part and symptoms, 2. colours, 3. name and cause. As depicted in Figure 12b body parts and symptoms are, in all groups, the most important criteria whereas name and cause are in general the least important criteria.

Differences between the groups. Categorizing dermatological disorders by their symptoms was most preferred by group 1 (L1/G) and group 3 (L2/G), the two groups that surfed on the German version. Group 2 (L2/G) and 4 (L2/E) preferred the body part criterion.

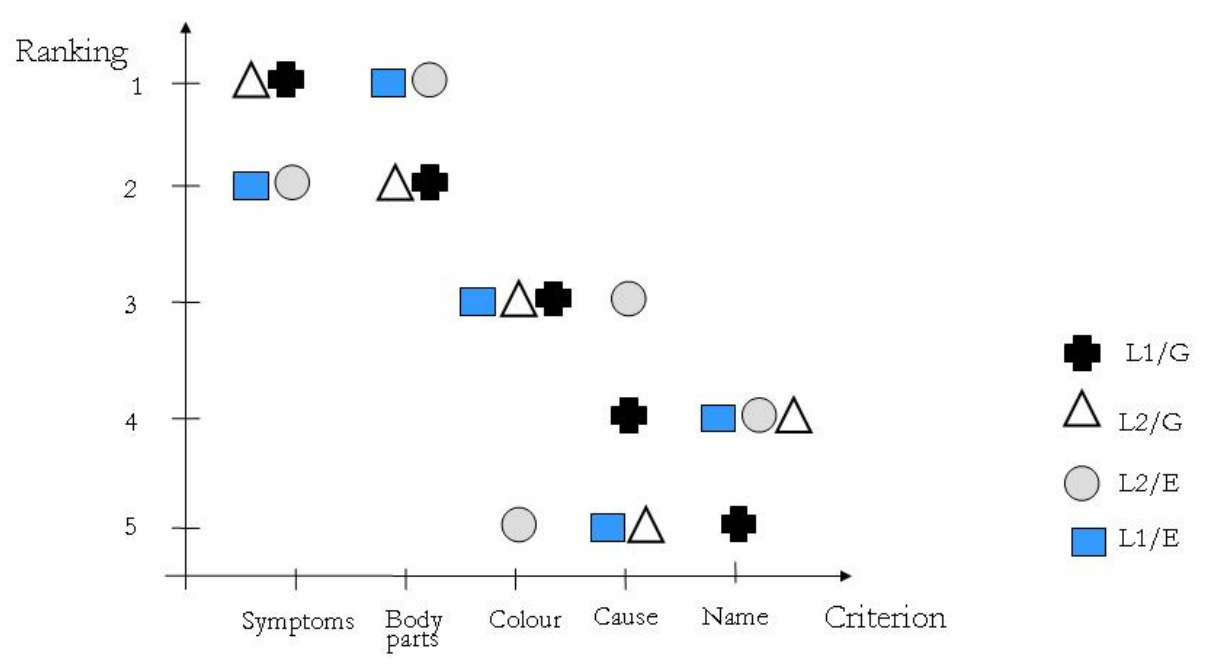
Group 3 (L2/G) represents an exception with regard to the cause criterion. It was the only group where cause was the third important criterion. This confirms results from pretests where the four Russian participants tested sorted the cards according to "social" criteria (contagious, insufficient skincare, etc.) rather than body parts or colours. However, the ranking does not confirm the importance of the criterion.

Figure 12a and 12b illustrate the results from card sorting and ranking. It is to be read as follows: The results of between-group comparisons are marked from left to right. On the left side are the groups that used a criterion the least, on the right side those groups that used it most often. The respective percentages of participants who used this criterion are marked above. We furthermore tested adjacent groups for significant differences in the use of the criteria for card sorting. Groups that are placed in the same rectangle do not show significant differences. Where significant differences between two adjacent groups exist, we put them in separated rectangles and marked the p-value. The corresponding result from the ranking of search options is marked below each criterion and each group. The average importance of the single criteria resulting from within-group comparisons is displayed from top to bottom: most used at the top, least used at the bottom.



* the display of the p-value indicates a significant difference between the adjacent groups

Figure 12a. Search Criteria Preferences Resulting from Card Sorting
(Between-groups Comparisons)



1=most preferred; 5=least preferred

Figure 12b. Search Criteria Preferences Resulting from Criteria Ranking (Within-groups Comparisons)

2.3.5.3.2 H16: Belief about Ease of Use

Answers concerning the *"ease of use"* of each search option were tested for reliability, which was deemed to be sufficient in each case (Cronbach's $\alpha > 0.8$ for each search option). Outcomes from reliability tests were confirmed by factor analysis, condensing variables for each case into one factor.

Results from the between-group analysis suggest that group 4 (L2/E) (and partly group 3, also an L2 group) finds all search options significantly easier to use than all other groups. Ignoring these ceiling effects, the within-group analysis reveals more valuable results, which are depicted in Figure 13.

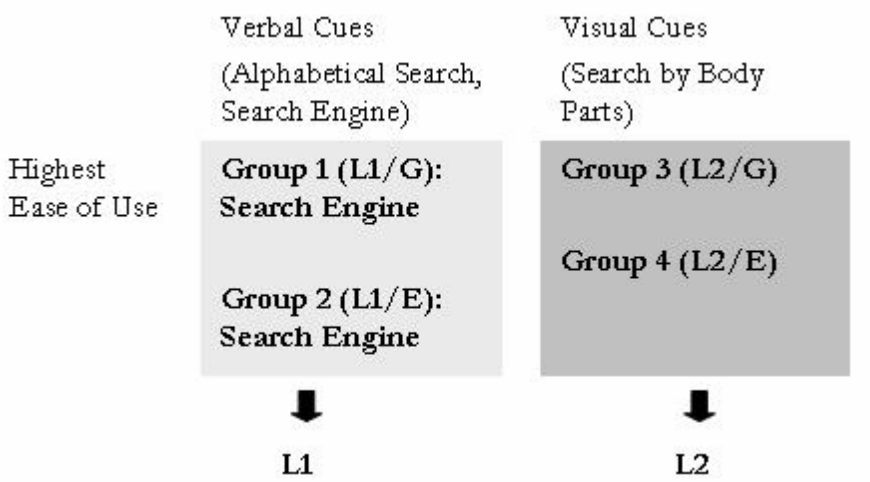


Figure 13. Belief about the Ease of Using Search Options with Verbal and Visual Cues Dependent on the Users' Linguistic Backgrounds

The figure shows that the search engine is considered to be the most useful search option within group 1 (L1/G) and group 2 (L1/E) (the two native speakers' groups) whereas participants from the two non-native speakers' groups, group 3 (L2/G) and group 4 (L2/E), considered the search by body part to be the most useful way of seeking information (see also Appendix A-2.4).

2.3.5.3.3 H17: Belief about Usefulness

Results are similar to those concerning the participants' beliefs about the ease of using the search options. Again, reliability is sufficient (Cronbach's $\alpha > 0.95$) for each search option and factor analyses resulted in one factor for each search option. Again, group 4 (L2/E) ranks all search options as more useful than any other group, so an examination of within-group results provides more useful insight.

The two L1 groups (group 1 and group 2) consider a search option that is based on verbal cues (alphabetical search, search engine) to be the most useful search option, whereas the two L2 groups (group 3 and group 4) find the search by body part, i.e. searching by visual cues, most useful. The outcomes are confirmed by the fact that the group 1 (L1/G) considers the search by body part to be the least useful search option. Figure 14 illustrates the results.

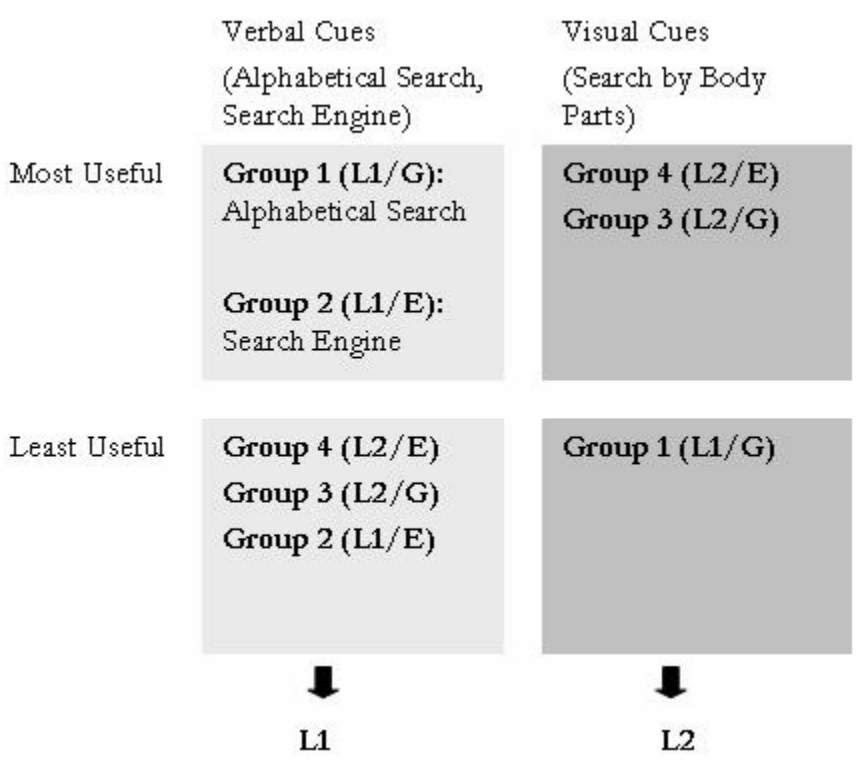


Figure 14. Belief about the Usefulness of Search Options with Verbal and Visual Search Cues, Dependent on the Users' Linguistic Backgrounds

2.3.5.3.4 H18: Attitude

Reliability analyses and factor analyses revealed the same outcomes as above (Cronbach's $\alpha = 0.942$), 1 factor for each search option.

Results from between-group analyses reveal hardly any significant difference between adjacent groups. Only the more negative attitude of group 4 (L2/E) towards the use of the colour criterion ($p=0.035$) and the more negative attitude of group 3 (L2/G) towards the use of alphabetical search ($p<0.001$) differ from the other groups (see Appendix A-2.4).

Furthermore, clear differences in the groups' attitudes towards visual cues (symptoms, body parts, colours) versus cues that require more extended linguistic and medical skills (alphabet, search engine), can be identified. However, these differences do not reflect results regarding the *"ease of use"* and *"usefulness"*. As shown in Figure 15, results from within-group analyses suggest that the group 4 (L2/E) exhibits, in general, a more positive attitude towards verbal cues and a more negative attitude towards visual cues, whereas it is the opposite case for group 1 (L1/G), group 2 (L1/E), and group 3 (L2/G). With the exception of group 4 (L2/E), the participants' attitudes therefore confirm the preference of visual cues over verbal cues obtained from card sorting and search cue ranking.

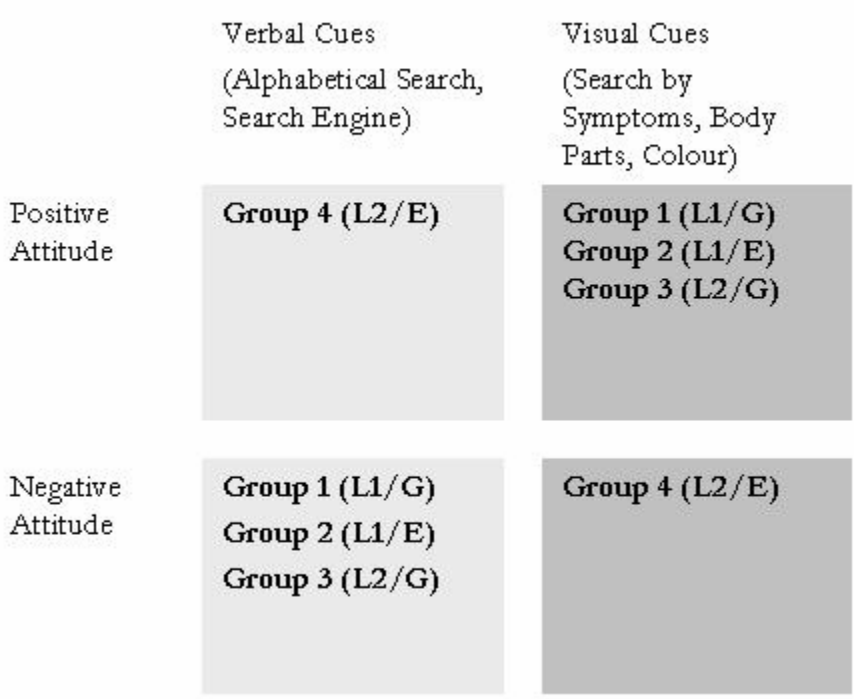


Figure 15. Attitudes towards Using Visual and Verbal Search Cues

2.3.5.3.5 H19: Use

The four groups investigated exhibited few significant differences with regard to their use of the search options offered (search engine, alphabetical search, search by body part). Only group 4 (L2/E) had (almost) significantly more people who used the search engine ($p=0.016$) and alphabetical search ($p=0.057$), and significantly fewer people using the search by body parts ($p=0.071$) - significantly less than the most extreme of the other three groups. The results confirm those from the analysis of the participants' attitudes. An examination of the using behaviour within the groups revealed that a search by body part is the most preferred, and an alphabetical search is the least preferred by each group. These results – favouring visual cues – confirmed outcomes from card sorting and search cue ranking (For more detailed results see Appendix A-2.4).

2.4 SUMMARIZING DISCUSSION AND IMPLICATIONS FOR WEBSITE DESIGN

2.4.1 The Impact of Culture

The results of the four studies presented broadly confirm the impact of cultural dimensions (as defined by Hofstede and Hall) on website navigation behaviour in general and search behaviour in particular. The large number of users examined extends previous research since the reconstruction, visualization and in particular measurement of navigation paths were so far applied only to a limited number of users and did not take cultural aspects into account. Our findings provide empirical

evidence for culturally differentiated navigation behaviour as a counterpart to the cultural differences in website design identified by Marcus and West Gould (2000).

The impact of culture regards three key aspects: time, space, and amount of information. The dimensions of Long-term Orientation and Monochronicity determine how time is valued and structured. As a consequence of time valuation, Short-term Orientation correlates with a preference for search engines and more pages visited per time unit. Furthermore, monochronic time structuring leads to more linear navigation patterns than polychronic time structuring. The dimension of Power Distance influences their relationship with hierarchical levels. Being used to deeper hierarchy structures, high Power Distance users prefer hyperlink searching more than low Power Distance users. The dimensions of context specificity and Uncertainty Avoidance relate to how much information is desired. Among these two, Uncertainty Avoidance is a rather complex dimension that requires distinguishing between behavioural and attitudinal outcomes. The dimension predicts that high uncertainty avoidant individuals feel the desire to collect as much information as available. Large amounts of information therefore create a feeling of anxiety and discomfort since exhaustive information gathering is not achievable. Hence, even though high Uncertainty Avoidance leads to intensive information collection, restricted information offers would reduce anxiety and increase comfort.

Opposite to what we expected, high uncertainty avoidant cultures show a higher preference for search engines than low uncertainty avoidant cultures. Inconclusive results were obtained with regard to the use of a content search. On the other hand, results from study 1 confirmed that users from high uncertainty avoidant cultures collect more information than those from low uncertainty avoidant cultures. Future research is needed to find explanations for this behaviour. In contrast, results regarding the dimension of context specificity are reflected in the preferences for search options and the users' need for information.

To sum up, there is evidence that cultural dimensions – in particular the amount of information needed and the perception of time and space – have an impact on the users' search behaviour. These differences are likely to be caused by inherent thinking patterns, which are determined by the users' cultural backgrounds.

Nevertheless, we also found a noteworthy commonality across all users: results suggest that there is a general preference for alphabetical searches over content searches, which in turn is preferred over search engines. This finding is in accordance with a number of other studies (e.g., Chimera and Shneiderman, 1994).

Website providers offering information to an international audience should take these results into consideration when designing search options and information access on their website. Analysis of the cultural indices involved suggests that monochronic cultures tend to be highly uncertainty avoidant, short-term oriented, and live in low context cultures (group 1). On the other hand, polychronic cultures tend to exhibit low Uncertainty Avoidance, are long-term oriented, and form high context cultures (group 2). These correlations allow us to cluster findings from our studies and resulting recommendations for website design into two groups (the left and the right column of the following table). It should especially be noted that each group contains complementary suggestions that can be realized on the same website. A division into two major design styles might therefore be considered, which leads us to the following recommendations for website design:

Table 9. Grouped Website Recommendations

	Group 1	**Group 2**
	Time	
Study 1	**Monochronic:** Information placed in linear order Links emphasizing hierarchical structure	**Polychronic:** Design supporting navigational freedom Cross-referencing links
Study 1 *Study 2*	**Short-term oriented:** Design minimizing time spent accessing information (e.g. search engine)	**Long-term oriented:** Design may ignore time aspects (and e.g. favour sophisticated pictures with longer download times Browsing schemes (content-organized links)
	Information	
Study 1	**High Uncertainty Avoidance:** Restricted information offer	**Low Uncertainty Avoidance:** Extensive exploration possibilities (e.g., many pages, pop-up windows, etc.)
Study 2	**Low Context:** Search engines	**High Context:** Browsing schemes (content-organized links)
	Space	
Study 2	**Low Power Distance** Search engine	**High Power Distance:** Hierarchically organized search options (content-organized links, alphabetical links)

2.4.2 The Impact of Language

Our studies also provided evidence of the impact of the users' native languages on website usage patterns, but the impact is not as discernable as that of culture. The relationship between the users' native language and navigation and information seeking behaviour can only be understood if the users' domain knowledge is taken into account. Due to the low number of participants in some sub-samples, results should be interpreted with caution. Confirmatory analyses based on a larger sample group are clearly needed.

The data showed a significant mediation effect of domain knowledge on the users' language skills. Terminological knowledge and cognitive effort are potential reasons for the higher percentage of physicians among non-native speakers than among the native speakers. Whether a search option is preferred by a physician or by a patient changes between non-native speakers and native speakers. Whereas in the L1 group all search options appear to be preferred by patients, results regarding the L2 group confirmed our predictions: The use of search engines and alphabetical searches are preferred by physicians, whereas patients exhibit a higher preference for content search.

L2 users with high domain knowledge do not differ from L1 users. Preferences for search options among this subgroup of the L2 user group do not seem to be determined by language. Hence, the hypothesized limiting effect of language was not confirmed for all non-native speakers. L2 users with high domain knowledge are the most likely to have the same success accessing information on a website as L1 users have.

Yet, in cases where a lower language proficiency level is combined with low domain knowledge, content-organized searches are clearly preferred over the use of search engines and alphabetical searches. Since both variables determine the degree of cognitive effort, language proficiency is exacerbated through low domain knowledge. We can conclude from these outcomes that search behaviour is sensitive to language proficiency only if combined with low domain knowledge. The area of language-sensitive search behaviour is consequently smaller than expected. This observation is depicted in Figure 16, which repeats and refines the expected structure shown in Figure 11.

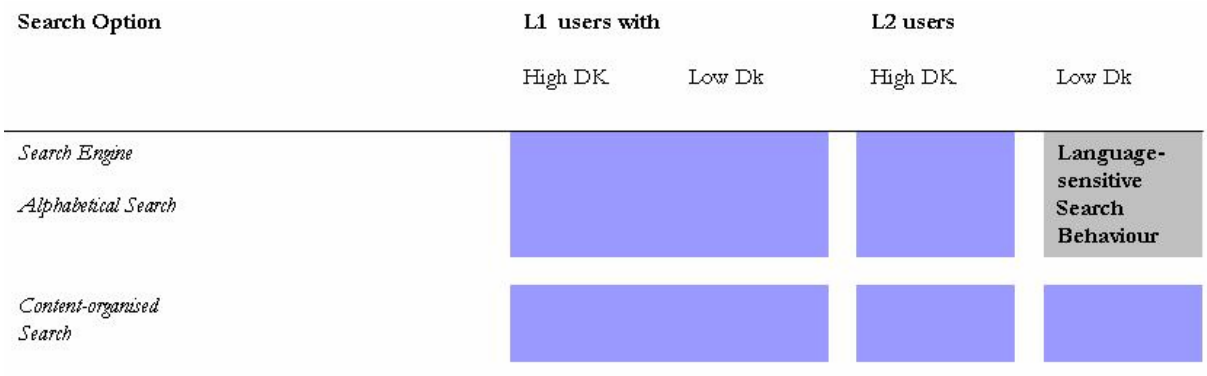


Figure 16. Overview of Observed Language-sensitive Search Behaviour

Similar mediating effects of domain knowledge on language proficiency have been shown for other areas, particularly with regard to L2 proficiency tests (McDaniel, et al., 2002; Steffenson, et al., 1979). Our results provide new insight into the same interaction in a different area, within the context of an interactive medium, the Internet. The outcome emphasizes the importance of the users' levels of prior knowledge – potentially resulting from their educational level - when the Internet is used as information source by L2 users. This should be taken into consideration when the goal of offering important public information on the Internet (e.g., health information) is to reach broadly, thereby helping undereducated parts of the population. In addition, the fact that L2 users' prior knowledge and education (or lack thereof) represent a barrier to obtaining further information leads to a deeper gap of knowledge level in underrepresented user groups. The data didn't provide evidence for a digital divide that separates majority language groups from underrepresented language groups. Rather, there appears to be a divide within underrepresented groups that segregates users having a higher education level from users of a lower education level.

It should also be noted that the average high language proficiency level of the users who participated in the survey is a first indicator that only L2 users above a certain proficiency level use non-native language Internet offerings. More details and results regarding the impact of language on website access, website satisfaction, and information flow are presented in Chapter 4.

Outcomes from study 4 provide further confirmation of the impact of language on the attitudinal level. Results suggest that information offered in the user's native language leads to higher perceived *"usefulness"* and *"ease of use"*. However, important basic preferences for information categorization appeared to be dominated here again by the users' domain knowledge: users with low medical knowledge preferred visual cues over verbal cues.

Finally, language as an overt characteristic of culture, as well as other cultural aspects, seem to influence secondary categorization preferences (i.e which type of non-visual cues). Results from empirical testing strongly suggest that within the visual cues the importance of *symptom* or *body part* as search cues differs between the analysed groups. Nonetheless, this exploratory part of our research needs to be extended by further analyses investigating the specific reasons for these differences.

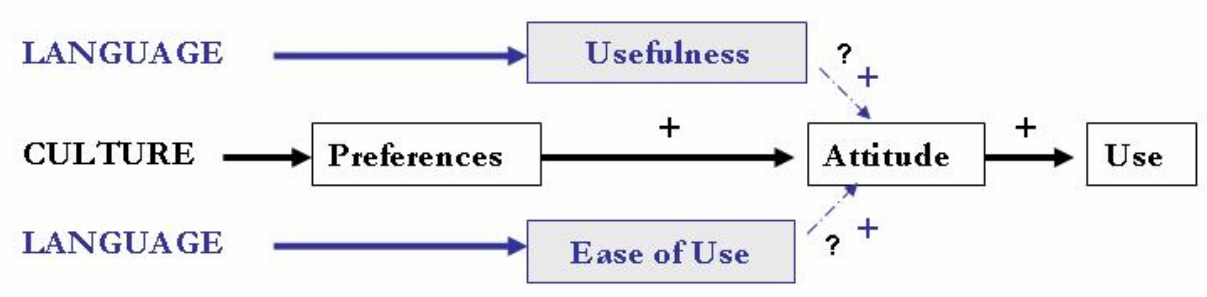


Figure 17. Impact of Culture and Linguistic Skills

Besides the consequences for a website's design in terms of structure, content, and language use, insight gained from our studies also have implications for usability testings whose generalizibility across cultural, linguistic, and educational borders is strongly questioned through these findings.

3. CHAPTER:

THE IMPACT OF CULTURE ON PERCEIVED RISK, PRODUCT VALUATION, AND ATTITUDES TOWARDS DATA DISCLOSURE

3.1 INTRODUCTION

It is important for providers of information systems to determine factors that affect the acceptance and use of their system. In the previous chapter's studies we investigated preferences for information presentation. We analysed patterns of behaviour based on the assumption that behavioural reactions are due to mental models and cognitive load. Lower cognitive load was assumed for more frequrent behaviours. Referring to the Technology Acceptance Model (Davis, 1993), the perceived *"ease of use"*[37] is only one of the two determinants. The other one is the *"perceived usefulness"*. After having investigated the impact of culture on cognitive burden, this chapter presents studies that analyse the impact of culture on preferences and needs for certain information but also for specific features.

Website providers offer products and services[38] on their sites, but whether they are used and how they are valued depends on a myriad of factors. Depending on the circumstances, Internet users might have to pay, disclose personal data, or simply spend time and cognitive effort in order to gain the product or service. It is a trade-off between personal resources and the value of a website's offer, where it is only worth the personal resource allocation if products and services are sufficiently valued.

Cost-benefit-trade-offs are a typical research subject in classic business-oriented information system studies. They seek knowledge about customers' willingness to pay for a product, but also look in a broader sense at their privacy concerns as well as possibilities to generate trust and loyalty (see below). However, little research has been done so far concerning the impact of culture and language.

---

[37] Since cognitive load/esae of use was only inferred from behaviour and not measured directly, we did not distinguish between objective cognitive load and perceived cognitive load.

[38] Marketing science usually considers a service as a product. Here we use the two words in a distinctive manner to emphasize the difference between products that are tangible (products) and those that are intangibile (services).

The question of which role culture plays in this utility framework is raised. At the same time, it brings up the problem of how previous findings in Information System research - usually obtained without taking users' cultural and linguistic backgrounds into account – can be generalized.

In order to obtain knowledge about these issues we conducted three studies in which we gathered data about cross-cultural differences with regard to the three aforementioned aspects: perceived risk, product valuation and attitudes towards data disclosure. These studies do not claim to provide exhaustive insight into culture related divergences by valuing websites and their services. They rather represent examples of how culture affects beliefs about their benefits.

Insight into the role of culture and language with regard to perceived benefits and costs of products and services is primarily useful for websites with a commercial goal: respective knowledge constitutes a competitive advantage, and may consequently help increase business volume or build loyalty. Nevertheless, websites that have a societal goal can also benefit from insight of this nature. Data about product and service needs helps attract a higher number of users and also helps explain the users' perceived barriers of accessing and using the website.

The main goal of health-related websites is to provide information to professionals and consumers. We assumed that users' evaluations of the website's benefits are based on two major issues: first, the compliance of the provided information with the users' needs (information content) and second, the compliance of the type of information presentation with the users' preferences.

The role of the presentation of information was extensively investigated in the previous chapter, where we provided evidence of the impact of culture, language, and domain knowledge on the users' preferences for navigational structures and information categorization.

In this chapter we focus on the role of the users' backgrounds with respect to their specific needs for information content, and restrict analyses to the impact of culture. The reason for this restriction is that we expected the impact of language on the need for content to be negligible. Accordingly, the first study (study 5) analyses cultural differences in risk perception (and hence information need) prior to the website visit. Subsequent risk reduction strategies on the website are examined. In a similar manner, the second study (study 6) aims to reveal culturally determined differences in the usefulness of various services. The second study differs from the first since it does not focus on the potential personal consequences of the lack of information, but emphasizes the preference perspective. Both studies investigate the benefit a user receives from visiting the website. The third study (study 7) emphasizes the cost aspect in its figurative sense, and investigates

the users' willingness to disclose data in this context[39]. Concerns of data disclosure are a particularly sensitive subject on health websites, where the advantage of anonymity has to be weighed against personalized service. Hence, data disclosure represents a risk for the users resulting from visiting the website.

In section 1.1.2 we distinguished between objective and subjective acceptance of information systems. Acceptance that is based on cognitive processes is basically related to differences in cognitive effort, and was the subject of investigation in chapter 2. Aspects examined in this chapter refer to acceptance based on attitudinal processes. Thinking patterns and habitual ways of thinking could lead to certain feelings of well-being, but also to discomfort. Due to this causal relationship, the users' cultural background is much more likely to influence subjective acceptance than language. Empirical studies therefore focus on the role of culture.

We introduce further aspects to broaden the concept of culture in this chapter. Product and service valuation are especially likely to depend on a user's current environmental settings, for example in terms of their economic situation and the country's health care system. These additional variables are even more justified in cases where the distinction between stable cultural values and current socio-economic settings is difficult (e.g., Tam, 2000).

3.2 LITERATURE REVIEW: RISK PERCEPTION, PRODUCT VALUATION, PRIVACY CONCERNS AND SUBJECTIVE ACCEPTANCE OF INFORMATION SYSTEMS

Product values, risk perception, privacy concerns, and cost-benefit trade-offs are traditionally investigated in business, economic, and psychological research. A countless number of theoretical and empirical studies dealing with these topics have been published over the last decades. For a survey see Brachinger and Weber (1997).

The vast majority of studies present results on the individual level without further aggregation. Nevertheless, there is reason to assume that the cultural background systematically affects individuals' perceptions of costs and benefits. Culture, as a determinant of personal values and "mental software"(Hofstede, 1980), through which the physical and social world is perceived, is

[39] Since we focus on health-related issues – with a rather societal goal -, we decided not to include the cost side issues of significant service remuneration in our investigations. Money-related aspects were however incorporated in study 6, yet in a different context and solely for methodological purposes.

prone to influence what is perceived as a benefit, what is perceived as a loss, and what is perceived to be the chances of either occurring.

Ko et al. (2004) showed, for example, that individuals' cultural backgrounds affect how aware they are of specific risks (see below) which results in a culturally determined need for information. On the other hand, the few cross-cultural studies that were conducted in this field found little influence of culture on overall risk behaviour. It can be assumed that insignificant results were obtained in even a larger number of analyses, but were not published due to their insignificance.

Predictions of online shopping behaviour (e.g., Jarvenpaa and Tractinsky, 1999) often refer to trust models that disregard potential specific needs. They only emphasize the users' perceptions of possible losses *caused* by the website offer (e.g. in terms of data disclosure, financial losses because product is not delivered or does not fulfil expected requirements). In Jarvenpaa's and Tractinsky's model the role of culture with regard to attitudinal variables is mixed. Culture was shown to directly affect trust, the antecedents of trust as well as perceived risk -which is an outcome of trust. On the other hand no link between culture and how much the purchased product was actually needed was found. Culture's overall impact was also found to be limited (see also 3.2.1).

Further cross-cultural examinations in the context of online shopping focus on culturally adapted design as antecedents of trust and website loyalty (e.g., Cyr, et al., 2004). Nielsen (2000) ascribes the link between website design elements and loyalty to positive cognitive user experiences.

In contrast to the abundance of predictive analyses of online shopping behaviour, analyses of the utility and usage behaviour of online information services are rare - in particular with respect to cross-cultural differences. In addition, existing insight was gained almost exclusively from laboratory experiments using students as test-subjects (see also Davison, et al., 2002). Our research therefore represents a contribution towards bridging this gap. We gathered data about (a) cross-cultural differences during the use of (b) online information services, and (c) primarily obtained reactive and non-reactive data collection from the audience of an existing website.

In our analysis of resulting risk reduction strategies we analyse participants originating from a large number of countries as suggested in section 1.3.3.2.1. The reason behind this is to diminish the potential impact of situational and national particularities, thereby limiting the risk of sampling bias.

In our first two studies we focussed on the benefit website users got from visiting the website. These investigations regarded differences in individual information need (and hence risk) before

visiting a website, and differences in the perceived utility of a website's products and services. Hereafter, *"cost"* perceptions of visiting a website are examined, namely in terms of data disclosure.

3.2.1 Culture and Perception of Risk through Information Deficit Prior to Visiting the Website (Benefit Side)

We refer to established concepts and measures that are widely used in marketing or psychological research. Cunningham (1967) classifies perceived risk as the importance of potential negative consequences as well as the probability that these occur. As a result of this initial risk's extent, individuals engage in various strategies to reduce their perceived risk. The influence of perceived risk on information search was investigated by a number of authors in the following years (Cox, 1967; Dowling and Staelin, 1994; Punj and Staelin, 1983; Srinivasan and Rathford, 1991; Sundaram and Ronald, 1998).

Jacoby and Kaplan (1972) introduced further facets of the multidimensional construct of perceived risk, differentiating between functional, financial, social, and psychological risk. Functional risk refers to the uncertainty of whether a product will perform as expected and financial risk denotes the price-quality relationship; social and psychological risks express the harm a product or service might cause to one's self-esteem or relationship with other people. In particular the latter two types of risk – social and psychological – are first indicators of the subjectivity of perceived risk. In fact, risk "…is an inherently subjective construct because what is considered a loss and what its significance and its chance of occurring is particular to the person concerned." (Brachinger and Weber, 1997).

Within the broader field of risk perception research, Gaenslen (1986) found cultural differences in the conceptualization of decision problems and their resolutions. Based on individuals' culturally determined world views, divergences occur in the perceived risk of nuclear power and in attitudes towards that technology, as shown by Peters & Slovic (1996). Weber et al. (1998) emphasize the importance of insight into cross-cultural differences of risk perception for understanding investment behaviour or creating bargaining solutions in negotiation. However, the authors also point out that possible reasons for national differences are twofold: long-standing cultural values as well as current situational circumstances (e.g. economic reward of investments) are determinants.

Investigations of cross-cultural differences in risk perception and risk behaviour on the Internet are rare. A first important contribution was provided by Jarvenpaa & Tractinsky (1999) who incorporated perceived risk in his cross-cultural model of consumer trust in an internet store. However, results from his sample of Australian, Israeli and Finnish students suggest little impact of

the users' cultural backgrounds on risk perception, nor on trust or willingness to buy. Australia is characterized by a much higher invidiualism score than Israel. Finland was added later. Its Individualism score is in between that of the two other countries. Similar results were obtained in a study comparing American and Korean online-shopping behaviour (Ko, et al., 2004). Countries were again chosen for their diverging Individualism level. Both cultural groups exhibited analogous degrees of perceived risk towards online shopping. Yet, findings from this study also provided evidence for culturally determined divergences regarding the significance of the various dimensions of risk. Whereas Korean participants systematically rated higher on social risk, Americans did so on time-related, financial, and psychological risks.

3.2.2 Culture & Perception of Risk through Information Disclosure on the Website (Cost Side)

Early studies in cross-cultural investigations of privacy concerns primarily aimed to *describe* differences in approaches and in attitudes towards privacy. Privacy was mainly investigated under the aspect of "How far can society intrude into a person's affairs" (Davies, 1996). In particular, divergences between Europe and the United States were emphasized: whereas in Europe privacy is perceived as a fundamental human right, it is considered part of contractual negotiation in the United States (Hessler and Freerks, 1995; Kirtley, 1999). Later on, studies focussed increasingly on investigating culture as an explanatory variable for these differences: In the vast majority of the studies subjects of investigations were privacy legislations in Europe (Harris, et al., 2003), the United States (Harris, et al., 2003) and Asia (Hann, et al., 2002; Ishikawa, 2000; Tam, 2000). Cultural values as defined by Hofstede, the contextual level of culture, communication patterns, and interaction norms were found to affect privacy perceptions (Rustemli and Kokdemir, 1993; Taylor, et al., 2001). Within these studies, Hofstede's cultural dimension of Individualism was mostly used to explain cultural differences of privacy perception. For example, Milberg, Smith and Burke (2000) found in their study of a cross-cultural sample from 19 countries that members of societies with a high Individualism index scores are more concerned about privacy. The authors showed furthermore that cultural values are mirrored in different regulatory approaches towards privacy. The finding of higher privacy concern within high Individualism societies was replicated by Smith (2001). However, a study conducted by Hann et al. (2002) was not able to show that the level of Individualism affects also the cost-benefit trade-off over information privacy.

Without referring explicitly to Hofstede's cultural system, Tam (2000) sees reasons for potential low awareness of privacy and the common violations of privacy in Asia in the habitual subservience to authorities, in the lack of private space, and in the uncommon notion of confidentiality and anonymity – the latter ones are a characteristic of low Individualism countries.

However, beside these classic cultural values, Tam also emphasizes the effect of the economic situations in most Asian countries and their ambition to boost their economies. Strict privacy legislation would constitute an obstacle to these economic ambitions. This shows that it is not a trivial task to determine the impact of culture on privacy concerns due to the difficult separation of culture from other processes, in particular socio-economic developments. The problematic separation is also due to the ambiguous causality between cultural values and socio-economic developments. As illustrated in section 1.1.3.2, evidence is provided in favour as well as against the existence of such correlations, and studies also indicate both directions of causal inter-relationships.

Moreover, in all of the mentioned studies culture was used as a given concept without questioning its appropriateness. In quantitative studies the concept of culture was generally based on Hofstede's cultural index scores, and consequently on Hofstede's view of culture. In non-quantitative studies, cultural research was for the most part based on general descriptions of national specifics.

3.3 EMPIRICAL WORK

3.3.1 Study 5: The Impact of Culture on Risk Perception and Risk Reduction Behaviour

3.3.1.1 Conceptual Framework and Hypotheses

The benefit or usefulness of a website's information is partly determined by the users' need for information. For the user, a lack of information may represent a certain risk - depending on the extent of the lack of information and its relevance. In the context of health information, knowledge about the negligibility of a disease might keep a patient from going to a doctor, saving them money and time. As a consequence, users may wish to gather information about a certain topic online and consequently reduce their risk.

The users' initial perceived risk is investigated with regard to financial, social, and time-related aspects – in line with Jacoby and Kaplan (1972 - see above). We relate to risk reduction as the extent to which a website (1) contributes towards the reduction in a user's perceived risk (regarding health related aspects). To measure the contribution of the information source (i.e. the website) for risk reduction, we relate to the perceived quality of information obtained, the contribution of a website's structural design in obtaining this information (search efficiency), as well as the users' overall satisfaction with the result.

The role of a website's structure/design was already examined in chapter 2. Yet those studies focussed on behavioural data without looking at success or satisfaction metrics. They did not provide any information about whether the website's given characteristics benefit one user group

more than the other. Does the website provide sufficient navigational choices in order to satisfy various user preferences? Do content and structure affect the users' attitudes towards the website? To test these variables we suggest a model that is described in the following section.

3.3.1.1.1 Endogenous Variables

The main structure of our model consists of three parts: the level of initial risk, risk reducing efforts, and risk-related outcomes of risk-reducing efforts. It is assumed that these three parts are causally linked to each other, from the first to the last. In addition, it is expected that the exogenous variables (basic and broadened aspects of culture) are determinants of each of these parts according to the following assumptions.

Level of Perceived Initial Risk. The level of perceived initial risk is the risk a user perceives prior to the website visit. As mentioned, financial, physical, psychological, and social risks are assessed – for the purpose of our study with respect to health related issues.

Risk Reduction Efforts. Risk reduction efforts result from evaluating the level of perceived initial risk. However, we also expect risk reducing efforts to be affected by exogenous variables. Risk Reduction Efforts are divided in a perceptual and a behavioural part: first, the users' expectations about the benefit of their search efforts and second, their actual behaviour, i.e. their actual search efforts.

Outcomes of Risk Reduction Efforts. Here again we differentiate between various elements to illustrate the different aspects of risk reduction behaviour: the perceived information quality (website content aspect), the perceived search efficiency (website structure aspect), and the overall satisfaction with the website in terms of perceived risk reduction.

3.3.1.1.2 Exogenous Variables

Culture. For the purpose of this analysis, the role of culture is investigated in terms of its stable (basic concept) and dynamic aspects (broadened concept). Two key issues are examined: the need for (medical) information and attitudes towards the medium, the Internet.

Uncertainty Avoidance has been classified in chapter 2 as a cultural dimension that describes individuals' need for information. The degree of a society's Uncertainty Avoidance illustrates a society's tolerance for uncertainty and ambiguity. High uncertainty avoidant cultures feel "threatened" by unknown or uncertain situations. Our study's outcomes (see chapter 2) indicate that the greater need for information of high uncertainty avoidant cultures results in more extensive

information collection (Kralisch, et al., 2005). In line with these findings we expect high uncertainty avoidant individuals to have a consistently higher amount of information need than low uncertainty avoidant individuals, influencing all three endogenous variables of the risk reduction model: the degree of the initial perceived risk, the search effort, as well as the search outcome.

H20a: Members of high uncertainty avoidant cultures perceive a higher initial risk than members of low uncertainty avoidant cultures.

H20b: Members of high uncertainty avoidant cultures engage in more search effort than members of low uncertainty avoidant cultures.

H20c: Members of high uncertainty avoidant cultures are less satisfied with the search outcome than members of low uncertainty avoidant cultures.

Whereas Uncertainty Avoidance determines an individual's general need for information, the dimension of Individualism is expected to play a role in this specific context. In line with Ko et al. (2004) we expect members of collectivistic countries to perceive a higher social risk than members of individualistic countries. In particular, easily visible or/and contagious diseases such as dermatological diseases (including sexual diseases) may lead to embarrassment and an unwillingness to talk about it among the patients concerned. These diseases are therefore likely to trigger differences in social risk perception. These issues of social risk are expected to be the main component of the perceived initial risk.

H20d: Members of collectivistic cultures perceive a higher initial risk than members of individualistic cultures.

The attitude towards the Internet is expected to be influenced by a culture's degree of Masculinity. The Internet is a relatively new medium. A culture with a high Masculinity index tends to foster long-standing traditions whereas opposite cultures show stronger tendencies to collapse traditional roles and values. Feminine cultures are less bound to traditional behaviour and should consequently be more open to receiving information besides the normal information paths (going to the doctor). Hence, feminine cultures are expected to expect higher benefits from using the website as an information source.

H20e: Members of feminine cultures expect a higher search benefit than members of masculine cultures.

Quality of Health Care System. The perceived quality of the health care system was chosen as an example of a socio-economic variable. The health care system available determines, for instance, how expensive health information from a physician is, how much time it takes to obtain medical information, and how reliable existing medical treatments are. A higher quality health care system reduces the importance of the health information offered on the website. Individuals from societies with a high quality health care system should therefore perceive a lower physical risk than individuals from societies with a health care system of lower quality.

H21a: Individuals from societies with a higher quality health care system perceive a lower initial risk than individuals from societies with a lower quality health care system.

Perceived information quality is compared to other available information and therefore depends on expectations formed by personal experiences, such that high experience leads to lower perception of information quality. Consequently, a better health care system is expected to diminish users' perceptions of information quality.

H21b: Individuals from societies with a higher quality health care system perceive lower information quality than individuals from societies with a lower quality health care system.

Web Experience and Domain Knowledge. Web experience has been shown to significantly influence risk perception (Jarvenpaa and Tractinsky, 1999; Ko, et al., 2004). Higher web experience leads to a lower perceived risk when shopping online. In line with these findings web experience is also assumed to have an impact on the expected benefit of search efforts.

H22a: The higher a user's web experience the higher the expected benefit of search efforts.

Here again it is not clear to which extent this variable is interrelated with the users' cultural background. A number of cultural characteristics could correspond with the peculiarities of the Internet better than others. As a result, the Internet might be used more by some cultures than others, leading to higher web experience. (e.g., more traditional, highly Masculinity cultures refer more to more traditional information sources, etc.).

Also, higher web experience provides the user with a large knowledge base for comparing websites. Similar to the impact of a health care system's quality, higher web experience should therefore lead to lower perceived quality of the website's information.

H22b: The higher a user's web experience the lower the perceived quality of a website's information.

In a parallel fashion, high domain knowledge increases the users' topic specific knowledge and is therefore expected to diminish perceived information quality.

H23: Users with high domain knowledge perceive a lower information quality than users with low domain knowledge.

Figure 18 depicts a simplified version of our proposed model. Endogenous variables (initial perceived risk, risk reduction effort and perceived risk reduction/satisfaction) are displayed within the grey box. The arrangement from left to right reflects the order and direction of impact, i.e. the initial perceived risk has an impact on risk reduction effort which in turn affects the perceived risk reduction/satisfaction. The exogenous variables of culture, quality of health care system, web experience and domain knowledge are displayed outside the box. The arrows indicate which endogenous variables are supposed to be influenced by the exogenous variables (e.g. the quality of health system affects the initial perceived risk and the perceived risk reduction/satisfaction.)

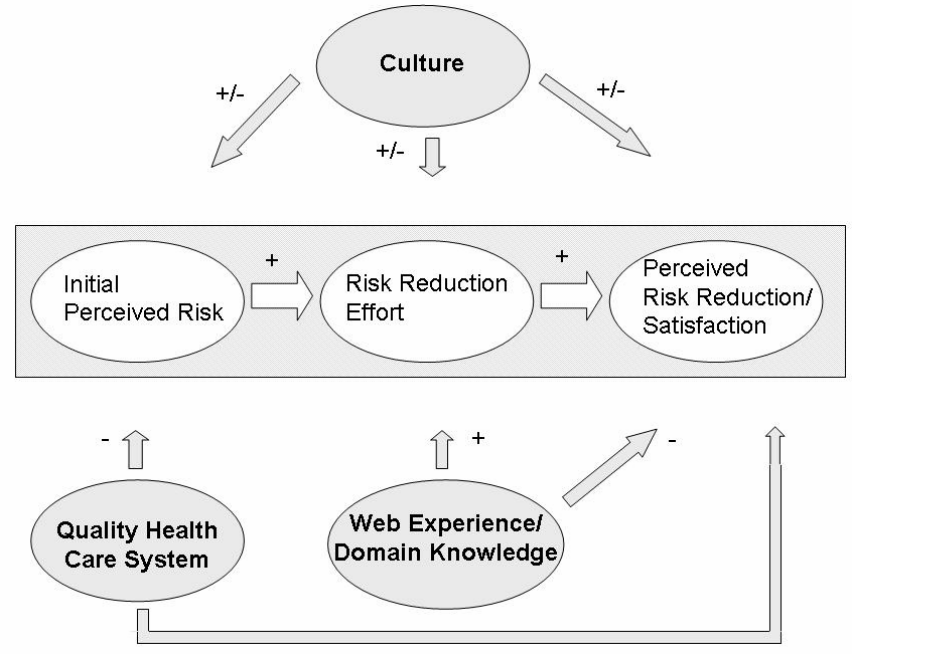


Figure 18. Simplified Proposed Risk Reduction Model

3.3.1.2 Method

3.3.1.2.1 Participants

Data was collected by means of a questionnaire posted on website A and the users' corresponding logfiles. Altogether, 183 website visitors from 39 different countries filled out the questionnaires.

3.3.1.2.2 Materials and Apparatus

The survey was conducted on website A between September and October 2004 as well as between May and August 2005. The questionnaire was linked through IP addresses and time stamps to the corresponding logfiles of the same period. Questions were presented to the users in their preferred website language. The questionnaire was developed in German and subsequently translated by native speakers into English, French, Spanish, and Portuguese. Details about the questionnaire design can be found in the Appendix A-3.1 .

3.3.1.2.3 Design: Measures

Cultural dimensions for Individualism (IND), Uncertainty Avoidance (UA), and Masculinity (MAS), were measured by country index scores taken from Hofstede (1980; 2001).

Perceived Risk (PR) was measured through five questions with answer scales proposed by Jacoby and Kaplan (1972). The authors multiplied the perceived probability of losses with the subjective relevance for each of the risk dimensions. Probability of losses and subjective relevance were measured with 7-point scales each ranging from *"rather unlikely"* to *"rather likely"* respectively *"hardly relevant"* to *"very relevant"*.

The Expected Benefits of Information Search (EBS) were measured with four 7-point scales indicating the degree of disagreement/agreement with the statements: *"It is worth consulting more than one information source for a medical problem", "I am sure to come to a better decision when I gather more information", "I know which treatment is the best for me when I extensively collect information", "There is much to lose if you don't inform yourself enough"* (Cronbach's a = .766).

Search Effort (SE) was measured by the number of page requests of each respondent as provided by logfile analysis.

The Perceived Quality of Information (PIQ) was measured with eleven 7-point semantic differential scales (reliable/unreliable, trustworthy/not trustworthy, accurate/inaccurate, objective/subjective, useful/useless, unimportant/important, relevant/irrelevant, outdated/up to

date, incomplete/complete, incomprehensible/comprehensible, easy to access/rather difficult to access) (Cronbach's *a* =.906). The respective positive and negative adjectives of each adjective pair were presented as mentioned here above. They were adapted for analysis purposes.

The Perceived Search Efficiency (PSE) was measured with three 7-point semantic differential scales (inefficient/efficient, rather time-consuming/rather quick, unsuccessful/successful) (Cronbach's *a*= .793).

The Perceived Risk Reduction/Search Success/Satisfaction (PRR) was measured with four 7-point semantic differential scales (unsatisfied/satisfied, disappointed/pleased, negative/positive, contented/discontented) (Cronbach's *a* = .848).

The Quality of the Health Care System (QHS) was measured with five 7-point scales indicating the degree of disagreement/agreement with the statements: *"The health care in my country is one of the best", "Everybody can afford health insurance.", "The health care system is socially fair.", "I can choose my health care.", "It is easy to organize a visit at the doctor's (making an appointment, getting there, etc.)"* (Cronbach's *a* = .803).

General Web Experience (EXP) was indicated by the frequency of internet use that was measured by a dummy variable that distinguished between users who used the internet at least once during the last week and users who did not.

Domain Knowledge (DOM) was measured by a dichotomous variable that distinguished between patients and medical professionals.

3.3.1.3 Results

Since the distributions of respondents over the countries turned out to be very unequal, the assumption of a normal distribution for our analysis focussing on cultural differences was violated, particularly for the cultural index variables. Hence, we aggregated the data over the 39 countries. The aggregated data constitutes the database for our analysis. For the continuous variables we used the mean as an aggregate measure; for dichotomous variables we used the percentage of respondents who indicated having a medical profession or using the internet at least once during the last week.

Not all cultural indexes are available for each country, which leads to varying sample sizes for the different analysis procedures. Because of the rather small sample, which is a common problem in these kinds of cross-cultural studies, we report significance levels up to $p < .1$ for our results.

Our assumptions were tested with several regression models incorporating country indexes as independent variables, and the variables referring to several stages of the information search and risk reduction process as dependent variables. Additionally, we added the aforementioned explanatory variables: website language proficiency, the quality of the health care system, general web experience, and domain knowledge.

According to our hypothesis, the following structural equations can be derived:

(1) $PR = b_0 + b_1IND + b_2UA + b_3MASC + b_4QHS + e_1$

(2) $EBS = b_0 + b_1MASC + b_2EXP + e_2$

(3) $SE = b_0 + b_1PR + b_2EBS + b_3UA + b_4LTO + e_3$

(4) $PIQ = b_0 + b_1\ QHS + b_2\ DOM + b_3EXP + e_4$

(5) $PSE = b_0 + b_1SE + b_2PIQ + b_3LTO + e_5$

(6) $PRR = b_0 + b_1PSE + b_2UA + e_6$

The equations are exactly identified since every equation includes at least one exogenous variable. Using Stata 9.0, we estimated the regression models. If necessary, we used 2SLS-regression in order to take endogenous independent variables into account, which may occur in equations (3), (5) and (6). Beforehand, we performed the Hausman simultaneity test on those equations in order to test for endogeneity (Gujarati, 2003, p. 670-672).

In equation (3), the residuals indicate that there is no endogeneity for EBS (t=1.03, p=.317). However, for PR we found endogeneity (t=-2.80, p=.012). Hence, we used MAS and QHS as best predictors for PR as instrumental variables.

In equation (5) SE and PIQ may be treated as endogenous variables. For SE, the residuals indicate no endogeneity (t=.98, p=.341). However, for PIQ, the residuals indicate endogeneity (t=2.94, p=.009). Hence, we used a 2SLS regression with DOM, EXP and QHS as instrumental variables for PIQ in equation (5).

Also, in equation (6) the test indicates endogeneity (t=3.06, p=.006). We used SE, PIQ, LTO, and WLP as instrumental variables for PSE in equation (6). Table 1 in Appendix A-3.2 summarizes the significant correlations.

According to our results none of our hypotheses that concerned the impact of culture was confirmed. This is either due to insignificant outcomes (hypotheses H20a, H20b, H20d) or to an impact that contradicts our predictions (H20e, H20c).

In contrast, we found evidence supporting the significant impact of the perceived quality of the health care system, web experience, and Domain Knowledge on the perception of risk and information quality. Only the impact of the perceived quality of the health care system on the perceived quality of the information provided is non significant.

3.3.2 Study 6: The Impact of Culture on Product Valuation

3.3.2.1 Conceptual Framework and Hypotheses

Study 6 complements study 7 by focussing on cultural differences in valuing the website's content and design features. Culture is employed here again as a broadened concept encompassing stable and dynamic aspects.

The outcomes of study 5 suggest that stable cultural values hardly affect the initial risk perception and hence the need for information, whereas dynamic aspects of culture (the quality of the health care system) do. Individual differences in web experience and domain knowledge predominantly affect the evaluation of the website's content and expectations when using the website.

In study 5 information need was derived from perceived risks and risk reduction behaviours. It regarded the users' personal consequences of information acquisition or lack of thereof. Study 6 proposes a new approach for investigating the need for information: we examine the users' evaluation of website design and content and their willingness to pay a minimal sum for the various features. This approach emphasizes the users' preferences in a website's offer without asking for the consequences. Thus, both approaches are similar in their research goal but differ in their focal points.

For the purpose of this second study, we distinguish between two main types of website products offered. The first group includes all features that enhance the content of the website through providing additional information. The second group consists of access features. They facilitate the

access to information, either because they make information access faster or because they facilitate the understanding of the information.

The investigation of access aspects and content aspects exhibit a number of similarities with Davis' distinction of *"usefulness"* and *"ease of use"* in his Technology acceptance model (Davis, 1993). *"Ease of use"* is mainly determined through the website's structure and information presentation. Chapter 2 revealed preferences in information categorization and navigation and search behaviour, which are determined by the users' cultural background, linguistic backgrounds and their domain knowledge. *"Usefulness"* can be regarded as the compliance of the website's content with the users' information needs. Nevertheless, the distinction between *"usefulness"* and *"ease of use"* is not always trivial since *"ease of use"* affects the degree of perceived *"usefulness"*: belief of easier use leads to a higher perceived usefulness.

Based on the principles of Ajzen and Fishbein's Theory of Reasoned Action (Ajzen and Fishbein, 1977; Ajzen and Fishbein, 1980), the TAM also predicts attitudes towards/satisfaction with the system and resulting behavioural actions. *"Ease of use"* and *"usefulness"* are antecedents of users' attitudes towards/ satisfaction with the system. Consequently, the investigation of the users' needs of websites' design and content features is also helpful as a predictor of the users' satisfaction with the website and their corresponding website use. Evers and Day (1997) linked culture as an external stimulant to the Technology Acceptance Model (see Appendix A-3.3). Findings from their study revealed culturally determined differences.

As illustrated in the previous study of this chapter, cultures are expected to differ in their needs for specific and general information, even though empirical results revealed a significant impact in only in two cases. A higher need for information should result in a higher need for additional content. Differences in information need that were elaborated upon in study 1 regard the cultural dimensions of Individualism, Uncertainty Avoidance, and Masculinity. Resulting from the argumentation and results obtained in study 5, it can be expected that collectivistic (low Individualism), high uncertainty avoidant, and feminine cultures (low Masculinity) would have a higher need for information and hence for content features than their counterparts.

Basic Concept of Culture. Regarding the impact of stable cultural values on the need of content features, we derive the following hypotheses.

H24a: Members of collectivistic cultures have a higher need for content features (on a health website) than members of individualistic cultures.

H24b: Members of high uncertainty avoidant cultures have a higher need for content features than members of low uncertainty avoidant cultures.

H24c: Members of masculine cultures (who visit the website) have a lower need for content features than members of feminine cultures.

Results from chapter 2 showed how cultures differ in the way they access information. From this point differences in need for certain access features can be inferred. It was shown that short term oriented cultures prefer search engines more than long term oriented cultures, and tend to spend less time per page. Hence, it can be inferred that short-term oriented cultures have a higher need for access features that allow them fast access to information.

This leads us to the following hypothesis:

H25: Members of short-term oriented cultures have a higher need for design features that allow them fast access to information than members of long-term oriented cultures.

Broadened Concepts of Culture (Socio-economic variables). In line with our findings about the relationship between the perceived quality of the health care system and need for information, the impact of a health care system's quality is assumed as follows:

H26: Individuals from societies with a lower perceived quality of health care system have a higher need for content features (on a health website) than individuals from societies with a higher perceived quality.

Broadened Concepts of Culture (Individual variables). Results from the previous study suggest an effect of the user's domain knowledge (see 3.3.1.3 and 3.4.1). We therefore propose the following hypotheses:

H27a: Users with high domain knowledge have a higher need for content features than users with low domain knowledge.

In line with outcomes from the second chapter, the impact of domain knowledge on the need for access features that facilitate understanding can also be expected.

H28b: Users with low domain knowledge (patients) have a higher need for access features facilitating content understanding than users with high domain knowledge (physicians).

3.3.2.2 Method

3.3.2.2.1 Participants

Data collection was conducted in two steps with two different sample groups.

The first part of the data (study A) was collected by means of a questionnaire. After data filtering, we worked with a sample set of 93 participants originating from 30 countries.

In the second study (study B) data was obtained in a laboratory session. The sample consisted of 92 international and German students. We excluded the data from those students with a multinational background (for reasons see 1.1.3.5). Furthermore, only the participants who spent at least 75% of their lives in their home countries were considered. After filtering the data through these criteria we worked with a sample of 84 students.

The majority of these students were between 22 and 27 years old, and had at least a fair knowledge of English. 60% of the participants were female, 40% male. Each participant was rewarded with a 9€ cinema coupon.

3.3.2.2.2 Materials and Apparatus

The questionnaire posted on website A in the beginning of 2005 (January – March). The questionnaire was presented to the users in their preferred website language: English, German, French, Spanish, or Portuguese. For the English version of the questionnaire see Appendix A-3.4.

The one hour computer session took place at a major German university in December 2004. Details about the questionnaire used for this laboratory session can be found in the Appendix A-3.5.

3.3.2.2.3 Design: Measures

Basic Concept of Culture. Cultural measures are based on information about the users' country of origin. Given the number of participants and their unequal distribution across cultural groups, cultural index scores were transformed into dichotomous variables with a cut-off point of 50 (see also Hofstede, 2001). We assigned each individual the corresponding cultural index scores.

Broadened Concept of Culture. The quality of health care service was rated by the participants of the website survey on a 5-point Likert scale. Website survey participants also had to identify themselves as patients or physicians. Patients were treated as users with low domain knowledge and

physicians as users with high domain knowledge. For the laboratory experiment every participant was considered to be part of the patient group. Further dynamic aspects of culture were not assessed (see Appendix A-3.4 and A-3.5).

Need for Website Features. Similar to study 3 in chapter 2, the need for website features was measured in each study through three measures.

(1) The website survey asked participants to describe the quality of information and information presentation with regard to the aforementioned content and access aspects (Measure 1 - for details see Appendix A-3.5). Survey participants had to rate the aspects by means of 6-point semantic differentials.

(2) Measure 2 asked participants to choose three out of five features, that they considered as the most important to be offered on the website. At that time none of the features were part of the website. These new features are *"enhanced pictures"*, *"comprehensive hyperlinks", "information about alternative treatments"*, an *"enhanced search function"*, and *"terminological support"*[40] . For technical reasons, it was not possible to integrate the feature of *"enhanced pictures"* into the experiment.

(3) Measure 3 assessed to which extent users would be willing to pay a minimal sum for each of the five additional features. Answers were rated on a 5-point Likert-scale (1-*"Never"*; 2-*"Unlikely"*, 3-*"I am unsure"*, 4-*"Possibly"*, 5-*"Definitely"*). In order to minimize the impact of the users' economic backgrounds, a non-specific term (*"a minimal sum"*) was used that adapts to each user's personal background. The term furthermore indicated a minimal financial engagement with supposedly little impact on the user's financial situation.

Since experiment participants were not actual users of the website and were therefore possibly less motivated, measure 1 asked for reasons why people in their home country would want to visit the website. As an equivalent to measure 3, experiment participants had to allocate an exact amount of money to each feature (Measure 3). In order to generate a situation of modest financial contribution, total money allocation was limited to 5€. This limitation was applied equalize the distribution of "available" money across all groups. The willingness-to-pay for a specific feature was measured through the amount of money allocated to that feature.

[40] For reasons of confidentiality we are not able to provide a more detailed description of the features.

Afterwards participants were asked to specify their reasons for their chosen way of allocating the money on a 6-point Likert scale (Measure 2). It should however be taken into account that the money-allocation question was much more authentic for the participants of the survey than for those of the experiment. We refrained from asking survey participants to allocate a specific amount of money for two reasons: first, the subject of the survey might make survey participants believe that they the website will become a remunerated service in the near future which might have elicited negative attitudes and feelings. Second, results from such a survey might be strongly biased, if participant assume that the amount of money to pay for the website is affected by their answers. Participants whose goal is to pay as little as possible would also allocate as little money as possible.

The users' willingness to pay a minimal sum for the website features, as a measure of their need for these features, was chosen for the following reason: In line with insight from decision and economic theories, willingness-to-pay represents a trade-off between monetary value invested by the user and the value gained from the investment. The higher the users' willingness-to-pay, the higher the need for the product. Yet, as discussed in a myriad of articles, *"willingness-to-pay"* is a complex variable that depends on a number of determinants (e.g., Smith, et al., 1999). It is therefore complemented by two other measures.

Control Variables. Within the website survey we controlled for tuition fees for higher education in the users' home countries (0=*"no tuition fees"*, 1=*"very low"*, 2=*"low"*, 3=*"average"*, 4=*"high"*, 5=*"very high"*). We furthermore assessed in the experiment the annual income available (ranges: *"less than 500€"*, *"501€-700€"*, *"701€-900€"*, *"901€-1200€"*, *"more than 1200€"*).

In addition, we controlled for the users' general beliefs about the value of pay Internet services in general, as well as their use of them. The need to control for this impact is grounded on insight from information seeking behaviour and decision theory: In the very likely case of incomplete market information or non-justified effort for value evaluation, consumers can rely on simplified heuristics or specific cues when making their decisions (Häubl and Trifts, 2000). These heuristics are most likely to be based on experiences with previous purchases of Internet services. It is assumed that the belief of a generally positive outcome when purchasing on the Internet leads to a higher willingness-to-pay.

3.3.2.2.4 Procedure

For the survey among the website's users, a questionnaire was presented to each user who chose to visit the previously announced new version of website A. Each user who clicked on a link to visit a

first version of the new website was directed to the survey. Visitors however had the option to skip the questionnaire and visit the new website version without answering the questions.

Within the laboratory experiment, participants had to surf on the same website in a language of their choice. In order to provide a motivation for surfing, participants were presented a task that required to find information on the website. Questions were asked in English. Details of the order of questions are illustrated in the Appendix A-3.5. In contrast to the website survey, participants were shown pictures of potential future features.

Data analyses were conducted for each data pool separately.

3.3.2.3 Results

3.3.2.3.1 Study 6a – Website Survey

The Kolmogorov-Smirnov-Test revealed that none of our independent variables complied with the requirements of normal distribution.

The control variable "tuition fees" was only in one case correlated with an independent variable, namely with Individualism (χ^2 -test: p=0.010). In line with Hofstede's findings about the correlation of Individualism and income (see also section 1.1.3.2), users from highly individualistic cultures pay more for tuition fees than those from low individualistic countries.

No correlation was found between Domain Knowledge and the stable cultural variables. However, in contrast to our previous study, three cultural dimensions showed significant correlations with the perception of the quality of the Health Care System. Individualism, Masculinity (positive correlations), and Long-term Orientation (negative correlations).

Interestingly enough, in our sample these three cultural dimensions also correlated among each other: low Individualism, low Masculinity and high Long-term Orientation seem to be strongly interrelated as well as their counterparts.

We assigned *"alternative treatment"*, *"hyperlinks"*, and *"pictures"* to the group of content adding features, and *"search function"* and *"terminological support"* to the group access facilitating features. The assignment of *"pictures"* was based on the fact that in the field of dermatology photos represent an additional source of information – due to the strong visual nature of this discipline – rather than a feature that supports the understanding of existing verbal information. These assignments were

confirmed by a factors analysis[41] that was performed on the features with regard to measure 2 and 3.

Further factor analyses were carried out with the goal to reduce the amount of investigated website features. Yet, items of measure 1 did not load on the five investigated features as expected (see Appendix A-3.6). We therefore calculated the impact of the independent variables on each single item through MANOVA analyses.

Due to the diverging availability of cultural index scores, MANOVA tests took all cultural dimensions into account but Long-term Orientation. This cultural dimension has the least amount of data available and would considerably reduce the validity of our outcomes. LTO was therefore tested together with the broadened aspects of culture in separate calculations. Data was weighted in order to consider the varying number of users per cultural group. Only significant outcomes are listed in the following section.

Measure 1: Shortcomings of Existing Website Features

The entire range of the 6-point semantic differential were used for the users' evaluation of content and design needs. Answers averaged out between 2.34 (boring) and 3.51 (too many medical terms) with a lower outcome equaling a more positive evaluation.

Table 10. Culture and Website Evaluation/Product Need (Measure 1)[42]

Cultural dimension	*Feature*	*Means*	*p (from Manova)*	*F*
Basic Aspects				
IND	Boring	Low: 2.58 High: 1.96	0.019	6.31
	Insufficient graphic support	Low: 3.17 High: 2.45	0.071	3.55
	Too many medical terms	Low: 3.64 High: 3.24	0.068	3.64
UA	Not informative	Low: 2.65 High: 2.05	0.003	10.95
	Not	Low: 2.33	0.085	3.21

[41] All of our factor analyses refer to principal component analyses. Varimax rotation was employed due to expected correlations between the items.

[42] High means stand for a low rating/ high need in this and the following tables.

	comprehensive	High: 2.38		
	Boring	Low: 2.19 High: 2.05	0.007	8.75
	Insufficient picture support	Low: 2.44 High: 2.12	0.054	4.08
MAS	Not informative	Low: 2.53 High: 2.13	0.030	5.29
	Not comprehensive	Low: 2.88 High: 2.11	0.024	5.78
	Insufficient picture support	Low:2.94 High:2.03	0.037	4.85
	Consumer unfriendly language	Low: 2.94 High: 2.65	0.014	6.98
	Hard to access	Low: 2.63 High: 2.05	0.037	4.85
Broadened Aspects				
Perceived quality of health care system	Too many medical terms	See Appendix A-3.6 Low: low High: high	0.062	2.57
	Hard to understand	See Appendix A-3.6 Low: low High: high	0.020	3.56
	Hard to access	See Appendix A-3.6 Low: low High: high	0.022	3.56
	Hard to find	See Appendix A-3.6 Low: low High: high	0.005	4.79
Domain Knowledge	Too many medical terms	Patient: 4.60 Physician: 3.38	0.014	7.01
	Hard to understand	Patient: 2.80 Physician: 2.64	0.030	5.26

Basic Concept of Culture. In accordance with our hypotheses H24a and H24c, users from collectivistic cultures and those from feminine cultures are less satisfied with the content of the website and have therefore a higher need for information. In contrast to what we formulated in our hypothesis H24b, low uncertainty avoidant users have a higher need for content than high uncertainty avoidant cultures. However, this outcome is in line with results from study 5, where the results regarding Uncertainty Avoidance also contradicted our hypothesis.

In addition, feminine cultures also rate the information access facilities lower with respect to information access and understanding.

Broadened Concept of Culture. As expected, patients have more trouble understanding the content, which can be interpreted as a need for features that facilitate content comprehension (H27b). Furthermore, the perceived quality of the health care system significantly affects the perception of access tools that allow easy access and understanding.

Measure 2: Preference for New Website Features

The preferences for website features from all users ranged from 0.13 (*"terminological support"*) to 0.69 (*"pictures"*), where 0 signifies *"not chosen"* and 1 *"chosen"*. The results represent the percentages of users who considered the respective feature as one of the three most important.

Table 11. Culture and Feature Preference (Measure 2)

Cultural dimension	*Feature*	*Means*	*p (from Manova)*	*F*
Basic Aspects				
UA	Hyperlinks	Low: 0.73 High: 0.50	0.082	3.23
Broadened Aspects				
Perceived quality of health care system	Pictures	See Appendix A-3.6 Low: low High: high	0.069	2.42
	Alternative Treatment	See Appendix A-3.6 Low: low High: high	0.045	2.75

Basic Concept of Culture. Outcomes confirmed the reduced need for content among high uncertainty avoidant cultures previously obtained, also contradicting hypothesis H24b.

Broadened Concept of Culture. The perception of the quality of the health care system leads to significant differences with regard to preferences for content adding features (H25).

Measure 3: Willingness-to-pay for New Website Features

Users' willingness-to-pay a minimal sum averaged out between 2.05 (*"hyperlinks"*) and 2.42 (*"pictures and search function"*), where 2 stands for *"unlikely"* and 3 for *"I am unsure"*.

Table 12. Culture and Money Allocation (Measure 3)

Cultural dimension	*Feature*	*Means*	*p (from Manova)*	*F*
Basic Aspects				
MAS	Terminological Support	Low: 2.56 High: 2.05	0.006	8.56
UA	Hyperlinks	Low: 2.27 High: 2.07	0.029	5.23
	Alternative Treatment	Low: 2.39 High: 2.05	0.067	3.61
	Terminological Support	Low: 2.60 High: 2.07	0.000	15.97
Broadened Aspects				
Perceived Quality of Health Care System	Alternative Treatment	See Appendix A-3.6	0.001	5.98
	Hyperlinks	See Appendix A-3.6	0.010	4.04
	Pictures	See Appendix A-3.6	0.006	4.49
	Terminological Support	See Appendix A-3.6	0.010	4.06

Basic Concept of Culture. Previous outcomes concerning the differences between high and low uncertainty avoidant users' need for information were confirmed by measure 3 (H24b). Low Masculinity cultures as well as low uncertainty avoidant cultures are more likely to pay for *"terminological support"* than their counterparts.

Broadened Concept of Culture. The table shows the significant impact of the quality of the health care system on a user's likelihood to pay money for content-adding features, as predicted by hypothesis H26 as well as with terminological support.

To sum up, results suggest that the perception of the quality of the health care systems and the cultural dimension of Uncertainty Avoidance are the variables with the most consistent impact on the need for content and access features. Yet, while hypothesis 26 regarding the health care system can be confirmed, hypothesis 24b concerning the impact of Uncertainty Avoidance is rejected. All measures show opposite outcomes and are therefore in line with outcomes from study 5.

We found further cultural differences in the need for content and access features that were not yet reflected by the users' willingness to pay for these features and/or their selection preferences. The differences found are however in line with our hypotheses (24a and 24c) and regard the higher need

for information for collectivistic and feminine cultures. No significant impact of Long-term Orientation was found.

In similar manner, measure 1 shows a significantly stronger need for understanding facilitators among the patient group (low domain knowledge) than among the physician group (high domain knowledge) (H27b). This outcome confirms hypothesis 27b. However, the difference between patients and physicians was not reflected in measure 2 and 3. In contrast to hypothesis 27b, no indicators that would confirm hypothesis 27a and the higher need for content among physicians compared to patients were found.

3.3.2.3.2 Study 6b: Experiment

Results from descriptive statistic calculations resemble those of the first study. No variable, construct or factor follows normal distribution (Kolmogorov-Smirnov-Test). The same correlations between the cultural dimensions as described in the previous study were revealed.

In contrast to study 5, we found significant correlations between the cultural dimensions and the control variable, the attitudes towards remunerated services. A factor analysis preceded the control for bivariate correlations, resulting in two factors.

For those who pay for online services, factor analysis revealed two main purposes: for private use *(MONEY_PRIVATE)* and for professional (business, scientific) use *(MONEY_PROFESSIONAL)*. For those who do not pay for remunerated services, three key reasons can be identified: first, the unjustified price *(NOMONEY_PRICE)*, second, the lack of facilities (services not known, no credit card) *(NOMONEY_FACILITIES)* and third, the lack of need for these services *(NOMONEY_NEED)*(for details see Appendix A-3.6).

Two cultural dimensions correlate with one or more of these factors. High uncertainty avoidant and short-term oriented users mentioned private and professional reasons more often as reasons for paying for online services than users from their cultural counterparts. Furthermore, users from high uncertainty avoidant cultures also mentioned the lack of facilities as reasons for not using remunerated services more often (see Appendix A-3.6) than users from low uncertainty avoidant cultures.

Due to the strong relationship between the cultural variables and the control variables, the users' attitudes towards the remunerated services was incorporated in subsequent analyses. However, for reasons of data reduction and presentation clarity, we integrate the users' attitudes towards

remunerated services into our analyses, but restricted detailed analyses into the basic aspects of culture.

Measure 1: Reasons for Website Visit

We obtained two factors from factor analysis (Prinicipal-Component Analysis, Varimax). Factor 1, encompassing the items time, cost, and negligibility of the disease can be summarized as time and cost reasons (NEED_TIMECOSTS). Factor 2 includes the need for alternative and general information and stands therefore for "information" reasons (NEED_INFO). Embarassement loaded equally on both factors.

Table 13. Culture and Reasons for Website Visit (Measure 1)

Cultural dimension	*Feature*	*Means*	*p (from ANOVA)*	*F*
Basic aspects				
UA	NEED_INFO	Low: -1.31 High: 0.11	0.086	3.01
Broadened Aspects				
MONEY_PRIVATE	NEED_INFO		0.068	3.50
NOMONEY_PRICE	NEED_INFO		0.032	4.91
	NEED_TIMECOSTS		0.059	3.78

This measure asked participants why people in their home country would visit this website. Results show that the need for information leads more often to significant differences than time and money concerns.

Basic Concept of Culture. High uncertainty avoidant cultures consider the need for information as reason to visit the website more often than low uncertainty avoidant cultures.

Control Variable. It was mainly found to have an impact on information need. Only those users whose price-consciousness is a reason for not using the Internet mentioned time and cost reasons more often those who are less price-conscious.

Measure 2: Reasons for Money Allocation

In order to obtain more than one factor from factor analysis, we reduced the eigenvalue-criterion to 0.90. Outcomes again confirm our division into content adding and access facilitating features. The first factor includes all items that describe aspects of easily accessing and understanding the

information (fast access, easy understanding, efficient use, easy website use, comprehensible explanations) (REASON_ACCESS). Items that describe aspects of information quality load on factor 2: complete information, objective information, credible information (REASON_INFO).

Table 14. Culture and Reasons for Money Allocation (Measure 2)

Cultural dimension	*Feature*	*Means*	*p (from Manova)*	*F*
Basic Aspects				
MAS	REASON_ACCESS	Low: 0.73 High: 0.50	0.015	6.51
Broadened Aspects				
MONEY_PROFESSIONAL	REASON_INFO		0.048	4.156

Basic Concept of Culture. Alone, Masculinity had a significant impact: a higher percentage of users from feminine culture than from high Masculinity cultures mentioned access features as a reason for the chosen money allocation.

Control Variable. The users' attitudes towards remunerated services had a significant impact on mentioning information aspects as a reason for money allocation.

Measure 3: Money Allocation

Factor Analysis roughly confirms our outcomes from study 5 since *"search filter"* and *"terminological support"* load on the same factor, whereas hyperlinks and alternative treatment show different outcomes. No logical relationship between the way money was allocated and the reasons for money allocation was found (see Appendix A-3.6).

Table 15. Culture and Money Allocation (Measure 3)

Cultural dimension	*Feature*	*Means*	*p (from Manova)*	*F*
Basic Aspects				
IND	Hyperlinks	Low: 0.54€ High: 0.47€	0.016	6.33
UA	Alternative Information	Low: 0.33€ High: 0.90€	0.069	3.49
Broadened Aspects				
NOMONEY_PRICE	Search Filters		0.061	3.70
NOMONEY_FACILITIES	Hyperlinks		0.017	6.167

Basic Concept of Culture. According to the results, users from low Individualism countries allocated significantly more money to *"hyperlinks"* than users from high Individualism countries. Furthermore, high uncertainty avoidant users showed a higher preference for "*alternative treatment information"* than low uncertainty avoidant users, through their way of allocating money.

Control Variable. The users' attitudes towards remunerated online services had a significant impact on how money was allocated towards *"search filter"* and *"hyperlinks"*.

To sum up, the significant impact of Uncertainty Avoidance and Individualism on the need for additional information was shown. Yet whereas outcomes regarding Individualism confirm previous correlations, results regarding Uncertainty Avoidance contradict previous outcomes. In addition, low Masculinity cultures prefer access-facilitating features more than high Masculinity cultures. Finally, users' attitudes towards remunerated services also had a significant impact.

3.3.3 Study 7: The Impact of Culture on Data Disclosure

3.3.3.1 Conceptual Framework and Hypotheses

Whereas the previously reported studies in this chapter emphasized the benefit side of visiting a website (i.e. how a website reduces risk through meeting the users' information needs), in this study we focus on the cost side. The investigated cost aspect is the disclosure of personal data. Evidence for the privacy-cost/benefit trade off is given by Stone & Stone (1990) who show that privacy perceptions are influenced by the purpose of the data disclosure in their study. Similar insight is provided by Kobsa and Teltzrow (2005) who found that the users' willingness to disclose personal data depends on the reasons why the personal data needs to be collected.

In contrast to the previous studies we only take the basic concept of culture into account. The primary reason for this approach is the difficult quantification of potential socio-economic determinants that are supposed to have an impact (e.g., privacy legislation). Previous studies that took these determinants into account were qualitative studies (e.g., Tam, 2000). This approach would also be justified if a strong and direct interrelationship between culture and privacy legislation is assumed (as suggested by Tam (2000)) that does not allow a separation between these two variables. Besides, privacy legislation is usually limited to the boundaries of a country. Hofstede's data aggregation to the national level, in this case, perfectly corresponds to the subject of investigation. Yet, Hofstede's definition of culture will have less explanatory value if recent economic developments are the reason for changes in privacy legislation (e.g. introduce an open

approach to privacy to boost the economy, or introduce a stricter approach to privacy to promote international trade with specific countries (Tam, 2000). Nevertheless, it can be argued that the possibility of enforcing these changes as well as the acceptance within a country's population depends still on traditional value sets.

So far, only the cultural dimension of Individualism has been taken into account as a potential explanatory variable. However, for the purpose of this study we furthermore propose to examine the impact of Power Distance and its interrelationship with Individualism. The question of to which extent these stable cultural values determine a user's willingness to disclose personal data on the website is raised.

3.3.3.1.1 The Cultural Dimension of Individualism

Individualism refers to the role of the individual within a group. A high Individualism score indicates that the interests of the individual prevail over the interests of the group. Their societies and governments place individual socio-economic interests over the group, maintain strong rights to privacy, and restrain the power of the state in the economy (Marcus and West Gould, 2000). In collectivistic cultures (low Individualism), governments place collective social-economic interests over the individual. They may invade private life, and usually dominate the economy (Marcus and West Gould, 2000).

The influence of Individualism on privacy concerns has been investigated on two levels: the individual and the societal. Results from Smith (2001) showed that individuals from societies with high Individualism scores exhibit more privacy concerns than individuals with low Individualism societies. Furthermore, Milberg and her colleagues (2000) showed that legislation in high Individualism countries foster the protection of private space more than regulations in low Individualism countries. Low Individualism countries (such as the European countries) have adopted a regulatory model where public officials enforce a comprehensive data protection law (www.gilc.org/ privacy/survey/intro.html).

An important exception, however, are countries of anglo-saxon tradition such as the United States or Australia, which have adopted a self-regulatory approach where data protection is considered a trade commodity.

3.3.3.1.2 The Cultural Dimension of Power Distance

We argue that beside the aspect of Individualism the dimension of Power Distance plays an important role with regard to privacy concerns and legislation. Power Distance refers to the extent to which less powerful members expect and accept unequal power distribution within a culture (Hofstede, 1991). Members of high power distant countries are consequently more accustomed to fulfil authorities' requirements. It can be therefore expected that collection of personal data is more accepted and more frequently performed in high power distant countries than in low power distant countries.

Clarke (Davison, et al., 2002) for instance describes Asian societies as cultures whose members are habituated to being subservient to authorities and hence to provide personal information. Tam (2000) emphasizes the strict censorship of Internet communication in China. These are signs of a strong power of authorities and strongly hierarchical structures, characteristics of Hofstede's cultural dimension of Power Distance. Indeed, Hofstede's index scores indicate high Power Distance for each investigated Asian country. Furthermore, as Tam points out, "only a handful of countries have actual legislation in compliance with guidelines of OECD" which would protect to a certain extent the privacy rights of individuals and consider them as fundamental human rights.

Moreover, in cases where high power distance correlates with low Individualism, intrusion of the private space is furthermore justified by the interests of the society prevailing over the interests of the individual. It is beyond the scope of this paper to argue whether or not the fostering of group interests over individual interests favours hierarchical structure and authority power. Interestingly, a revision of Hofstede's cultural index scores reveals that Power Distance and Individualism show a strong and significant negative correlation ($r=0.665$ (Pearson), $p<0.000$). The vast majority of the countries belong either to the group of high Power Distance and low Individualism index scores or the group with a low Power Distance and high Individualism index score. In the first group, countries are predominantly located in South and Central America, Asia, Africa, Arab countries, and in parts Southern Europe, in the second we find mainly European countries and non-European countries of Anglo-Saxon tradition (Hofstede, 2001, p.54).

Extending previous research, we therefore argue that for the purpose of investigating privacy issues, the impact of Individualism can only be evaluated if the role of Power Distance is taken into account, too.

3.3.3.1.3 The Model

In our study the external stimulus investigated is the user's cultural background. It is expected to affect the user's attitudes towards data disclosure, which in turn is assumed to affect the willingness to disclose personal data.

The expected direct and indirect impact of culture is expressed in the following hypothesis:

H28: Members of societies with a high Individualism index score and a low Power Distance score perceive a significantly higher hazard in disclosing their personal data than members of societies with a low Individualism index score and a high Power Distance score.

H29: Individuals with positive (negative) attitudes towards data disclosure are more (less) willing to disclose personal data.

We will also use the data to test the expected non-significance of the direct effect of culture on the users' willingness to disclose personal information.

3.3.3.2 Method

3.3.3.2.1 Participants

Data was obtained from a questionnaire. 173 website users answered the questionnaire. After manual data revision and the selection of data sets for which Hofstede's index scores are available, we analysed answers from 166 participants originating from 26 countries. Only those participants who were still residing in their country of origin were taken into account.

3.3.3.2.2 Materials and Apparatus

The data for this study were obtained in the same survey based data collection as presented in 3.3.2.2.2. The questionnaire was posted on website A in the beginning of 2005 (see Appendix A-3.7). The questionnaire was presented to the users in their preferred website language: English, German, French, Spanish, or Portuguese. Each user who clicked on a link to visit a first version of the new website was directed to the survey. Visitors however had the option to skip the questionnaire.

3.3.3.2.3 Design: Measures

Culture. According to their countries of origin and residence, we assigned survey participants to the following four different cultural groups: Group 1: high Individualism and low Power Distance,

Group 2: low Individualism and high Power Distance, Group 3: high Individualism and high Power Distance, Group 4: low Individualism and low Power Distance.

82 users from the following countries were classified into group 1: Austria, Denmark, Germany, Ireland, Netherlands, Norway, South Africa, Switzerland, UK, and USA. Group 2 includes 48 users from Brazil, Egypt, Greece, India, Nigeria, Pakistan, Portugal, Sudan, Syria, Tunisia, Turkey, and Yugoslavia. 36 users from Belgium, France, Italy, and Spain constitute group 3. No users were assigned to group 4 (Group distribution is in accordance with Hofstede (2001). Since we opposed the two most extreme groups in this analysis, group 1 and group 2, users from group 3 were excluded from further analyses.

Measures are the following:

Attitude towards Disclosing Personal Data. A 5-point anchored Likert scale was employed to ask users about their attitude towards data disclosure. Where necessary, items were recoded. The answers were coded as follows: 1-*"strongly disagree"*, 2-*"disagree"*, 3-*"neither agree nor disagree*", 4-*"agree",* 5-*"strongly agree"*.

Willingness to Disclose Personal Data. Users had to rate on a 5-point anchored Likert scale how comfortable they would feel to provide a number of specified personal information for registering for the new website version. Answers were 1-*"I would never provide this information"*, 2-*"uncomfortable"*, 3-*"doubtful"*, 4-*"comfortable",* 5-*"completely comfortable"*.

3.3.3.3 Results

Attitude towards Data Disclosure. The reliability test resulted in a coefficient of Cronbach's α=0.630, which is slightly below a sufficient level. Subsequent factor analysis confirmed this outcome by revealing two factors: the first one (DATA_NEG) encompasses the three items that focus on the potential negative aspect of data disclosure (data abuse, time). The two factors that emphasize positive aspects of data disclosure load on factor 2 (customization) (DATA_POS). However, it should be noted that both factors are coded so that a high value represents a negative attitude and a low value a positive attitude towards data disclosure.

Willingness to Disclose Personal Data. In a similar manner, factor analysis (principal component analysis, varimax rotation) was performed on the items asking for the users' willingness to disclose specific personal data. The eigenvalue > 1 criterion resulted in a two-factor solution. Items of the second data group are: *"provide email-address"* and *"provide information"* about the current

health or medical history; those of the first group are: *"provide specific information about medical profession", "provide information about medical knowledge"*, and *"provide information about country of origin"*. The item grouping suggests that items of the first group are highly privacy sensitive data whereas those of the second group are less privacy sensitive.

A control of the percentages of patients and medical professionals revealed a significantly higher percentage of medical professionals in group 2 compared to group 1 (χ^2-test, $p<0.000$).

In order to test the impact of culture we performed bivariate correlation analyses with weighted data to reflect the different number of users.

Culture and Attitude towards Data Disclosure (H28). An (almost) significant impact of the users' cultural background on the perceived hazard of disclosing personal data was found in the analysis. Having coded group 1 with the dummy variable 1 and group 2 with 2, the Spearman-correlation coefficient indicates that users with a low Power Distance and high Individualism background show higher privacy concerns than users from the second group. The correlation is slightly stronger for the first factor (DATA_NEG) that emphasizes the negative consequences of data disclosure (DATA_NEG: $r=0.180$, $p=0.053$; DATA_POS $r=0.170$, $p=0.069$).

No significant correlations were found, when each cultural dimension's impact was tested separately (IND-DATA_NEG: $p=0.892$; IND-DATA_POS: $p=0.324$; PD-DATA_NEG: $p=0.413$; PD-DATA_POS: $p=0.701$.

Hypothesis 28 was therefore confirmed.

Attitude towards Data Disclosure and Willingness to Disclose Personal Data (H29). Correlation analyses that controlled for the impact of culture were carried out. Among the four factors, one significant correlation was found: a more positive attitude towards data disclosure leads to a higher willingness to disclose what we called non-privacy sensitive data ($r=0.222$; $p=0.039$). As illustrated in table 16, no other significant correlation was found.

Table 16. Correlations between Attitudes towards Data Disclosure and Willingness to Disclose Data

Control Variables			*Privacy_Contra*	*Privacy_Pro*
Zscore(Cult_group)	data_privacyinsensitive	Correlation	,222	,123
		Significance (2-tailed)	,039	,257
		Df	85	85
	data_privacysensitive	Correlation	,139	,128
		Significance (2-tailed)	,199	,236
		Df	85	85
	Privacy_Contra	Correlation	1,000	-,028
		Significance (2-tailed)	.	,794
		Df	0	85
	Privacy_Pro	Correlation	-,028	1,000
		Significance (2-tailed)	,794	.
		Df	85	0

No direct impact of culture on the willingness to disclose personal data was found (see Appendix A-3.7). Thus, the results confirm hypothesis 29.

3.4 SUMMARIZING DISCUSSION ON THE ROLE OF CULTURE WITH REGARD TO PERCEIVED RISK, PRODUCT VALUATION, AND ATTITUDES TOWARDS DATA DISCLOSURE

A summarizing discussion of the outcomes presented can only reflect a simplified overview of the results obtained. Especially in study 5 and 6, the large number of measures, items, and partly consisting outcomes would require a detailed result presentation. Nevertheless, commonalities across the studies can be found, but they need to be interpreted with caution.

3.4.1 Benefit-Side

Study 5, 6a and 6b investigated aspects of the impact of culture on information need and how these are met with the website's content and design offer.

In study 5 we found only weak support for the risk reduction model. Only perceived information quality, search efforts and perceived search efficiency show results that correspond to the model, all other paths turned out to be not significant. An explanation of these results is provided by a more thorough look at the central construct of perceived risk. Gemünden (1985) concluded that models of risk reduction strategies appear to be particularly valid for high-involvement goods. In contrast, lower levels of product risk do not seem to trigger information search as a risk reduction strategy. It can be concluded that the health information service – as analysed in our case – is perceived as a low risk product.

Impact of Stable Aspects of Culture (Basic Concept). Study 5 only found a small impact of culture in terms of cultural dimensions as defined by Hofstede. The impact that was found contradicted our hypotheses: after gathering information on the website, high uncertainty avoidant cultures showed a higher satisfaction from which a lower need for information can be cautiously inferred. The variable Masculinity had an impact on perceived risk and expected benefits of search. Individuals from high Masculinity societies perceived a lower risk than members from more feminine cultures.

We consider this rather homogeneous risk perception and risk reduction behaviour across the cultural groups as a further indicator of a low-involvement. Low levels of product risk can trigger only little differences between individuals. Perceived risk and risk reduction strategy are consequently rather similar.

Results from study 6a and 6b confirm the suggested impact of Uncertainty Avoidance on information need: users from low uncertainty avoidant cultures exhibit a higher need for additional information than users from low uncertainty avoidant cultures. Alone, measure 3 in study 6b revealed an opposite impact: high uncertainty avoidant users mentioned information need as a reason for visiting the website more often than low uncertainty avoidant cultures.

Uncertainty Avoidance: Uncertainty Avoidance was identified as the cultural dimension that determines the users' need for information in chapter 2. Results strongly suggest that this role is confirmed with regard to attitudinal variables.

Nevertheless, most findings contradicted our predictions: lower uncertainty avoidant cultures appeared to exhibit a higher need for information that does not seem to correspond with the meaning of the dimension. In fact, in line with other studies' outcomes, Uncertainty Avoidance appears to be a complex cultural construct whose impact is difficult to predict (e.g., Kralisch and

Berendt, 2004). On the one hand, high uncertainty avoidant individuals tend to collect more information than low uncertainty avoidant individuals (Kralisch, et al., 2005). On the other hand, high uncertainty avoidant website users prefer a website design with limited choices (Marcus and West Gould, 2000). Our outcomes raise the question of whether information need is evaluated differently depending on whether *additional* information is desired or whether *access to existing* information is desired. According to our results, additional information is wanted by low uncertainty avoidant users, whereas access to existing information is preferred by high uncertainty avoidant users. Such a distinction of information need would be in line with the simultaneous preference for limited choices and extensive information collection of high uncertainty avoidant users.

Individualism: Isolated evidence for a significantly higher information need from users from collectivistic societies was found. In particular, the higher preference for *"hyperlinks"* is in line with our argumentation of higher website connectivity of low individualistic cultures, as suggested in (Kralisch and Mandl, 2005).

Masculinity: In a similar manner, results from study 5, regarding the higher need for information for users from feminine cultures were sporadically confirmed. The lower risk of high Masculinity cultures contradicts the predicted negative attitude towards the new medium of the Internet. Yet, taking into account the low-involvement assumption, it can be inferred that members of highly masculine societies use the Internet as an information source only in low-involvement situations where the risk is rather negligible. Two significant correlations also exhibited the higher preference of feminine cultures for features that facilitate the access and understanding of information.

No impact was found with respect to the impact of long-term orientation.

The Impact of Dynamic Aspects of Culture (Health Care System and Domain Knowledge)**.** According to the outcomes from study 5, socio-economic and individual aspects of the broadened concept of culture have a stronger impact on users' need for information than the classic cultural values. The higher perceived quality of the health care system led to a lower information need; higher web experience and domain knowledge correlated with a lower rating of the perceived quality of the information.

The impact of domain knowledge was also investigated in study 6a, where evidence was found for a higher need among patients for features that facilitate the understanding of information.

Furthermore, the significant impact of the health care system on the need for additional information as well as for features for fast and easy access to information was confirmed by all measures in study 6b. Similar to the impact of domain knowledge from study 5, outcomes suggest that a higher quality of the Health Care System leads to lower ratings.

Besides, study 7 found a strong impact of the control variable - the users' general attitudes towards remunerated online services which emphasizes the role of individuals' knowledge and habits.

The role of the users' attitudes towards remunerated online services also point out the problem of causality in cross-cultural research. The difference between control variables and variables that are incorporated into the broadened concept of culture lie in the research design and researcher's development of the conceptual framework. Problematic bias in the sample set was indicated by correlation between the perceived quality of the health care system and cultural dimensions that were present in the second study but not in the first.

3.4.2 Cost-Side

Study 7 differed from the first two studies since it examined the negative aspects of visiting a website, with regard to the disclosure of personal data. We also focussed only on the impact of the stable aspects of culture. The study provided evidence for the interrelationship between low Individualism and high Power Distance and vice versa, and their significant impact on the users' attitudes towards data disclosure. Users from highly individualistic and low Power Distance cultures exhibit a more negative attitude towards data disclosure than users with opposite cultural characteristics. This outcome confirmed hypothesis 28. In turn, a more negative attitude led to a lower willingness to disclose personal data, which was in line with hypothesis 29. Yet this correlation only applie to what we called *"non privacy sensitive data",* and not to *"privacy sensitive data"* (email-address, medical history/current health status).

To sum up, general conclusions should be drawn only with caution. Further research is hereby clearly encouraged. Results suggested that searching for health information was – in contrast to most cases of online shopping – a low risk situation. As such, classic models of risk reduction strategies can hardly be applied.

Stable cultural variables seem to have less impact than expected. These findings are in line with other studies (e.g., Jarvenpaa and Tractinsky, 1999; Ko, et al., 2004) where small impact of culture on risk perception and hence on information was found. They also suggest that motives and outcomes of information search on the Internet seem to be rather independent of cultural values,

which may be due to cosmopolitan users that are less determined by culture-bound mental systems. The most important cultural dimension appears to be Uncertainty Avoidance. This, in line with our outcomes from study 6, was identified as the dimension that determines the amount of information needed. The limited general explanatory value of culture is not specific to our study but rather common in cross-cultural research (Karahanna, et al., 2002).

On the other hand, results suggest a consistent impact of broadened aspects of culture such as the impact of the health care system (as example of socio-economic variable) and the medical knowledge. The impact of domain knowledge confirms results from chapter 2. However, the selection of variables that are considered part of the broadened concept of culture is strongly influenced by the research design.

4. CHAPTER:

THE ROLE OF LANGUAGE IN THE CONTEXT OF THE WORLD WIDE WEB

4.1 INTRODUCTION

In study 3 (presented in chapter 2) we compared the behaviour of L1 users and L2 users. The server-centric approach restricted the investigation to those users who accessed the website. This raises the question of the degree to which the access of the website was determined by language: either because only non-native speakers above a certain language proficiency level visited the website or because language determined how the website was linked to others. In order to answer the question, the role of language as a potential barrier to access is examined in study 8 in this chapter.

In our theoretical discussion it was stated that non-native language navigation leads to a higher cognitive burden (see section 1.2.2.1.1). This statement was partly confirmed by results from study 3 (see section 2.3.4.3.2). Study 8 aims to add information linking cognitive burden to user satisfaction. Since the acceptance of a language is also expected to depend on the language offer of other providers (see section 1.2.2.2.1), satisfaction is also linked to the perceived amount of native language information online in the same study, i.e. to the perceived amount of native language alternatives.

The World Wide Web is a network with websites as nodes and hyperlinks representing the edges between the nodes. Websites are therefore mutually dependent entities that constitute a web-system (Park and Thelwall, 2003). This is particularly true with respect to their informational value. The structure of the WWW, or the way websites are linked to each other, determines the information flow and knowledge generation to a large extent.

As a network, the Internet can be investigated from a technical perspective as well as a social perspective. An example of technical investigation is the calculation of characteristic metrics such as the centrality or connectivity of a website. Algorithms that are based on these metrics are used, for example, to rank pages for displaying search results.

These algorithms are based on the assumption that the distribution of hyperlinks represents similar principles on the entire Web. Suspicion that language and culture possibly bias link setting and link

using behaviour is raised. In fact, initial studies indicate the impact of culture: for example, Baeza-Yates and Poblete (2003) found national differences in hyperlink distribution.

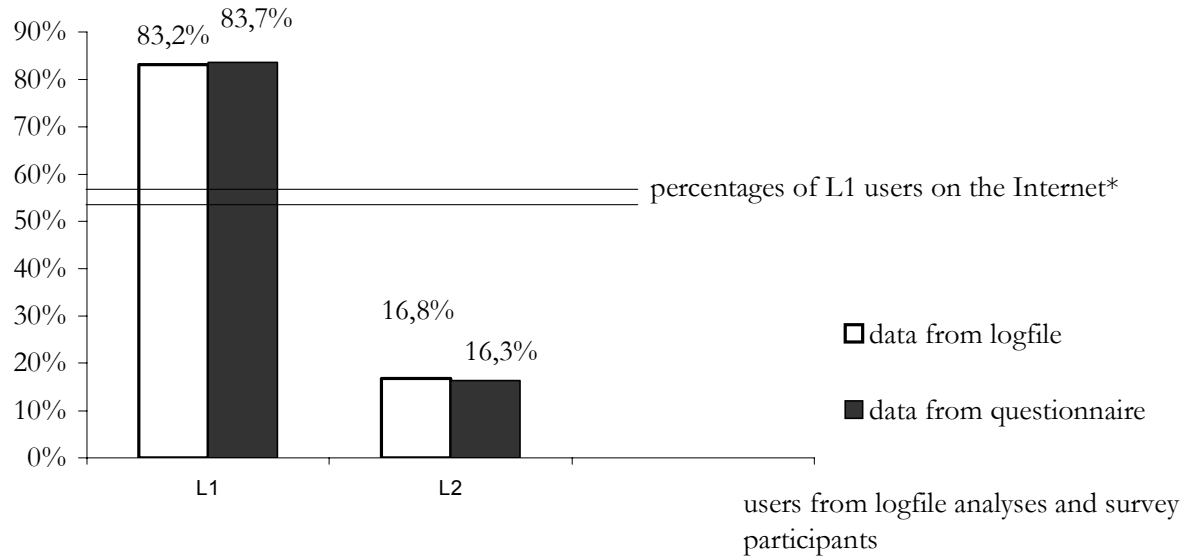
Information flow can be described in terms of its intensity as well as with respect to its direction. Culture and language are expected to affect both the number of links established between websites as well as which combination of websites are linked to each other. The term information flow is usually used in the context of hyperlink distribution, i.e. how hyperlinks are set to link webpages. However, we also consider it useful to examine the "flow of Internet *users*" within this framework, by analysing their navigation patterns through a website.

Results from chapter 2 suggested that a higher percentage of physicians within the L2 user group accessed the website than within the L1 group. The divergence was explained with differences in domain knowledge and language related cognitive effort.

Another outcome of the two studies in chapter 2 is the fact that – independent of domain knowledge - the ratio of L1 users accessing the website considerably exceeds that of L2 users.

Among the 277,809 sessions examined in the logfile-based study, we were able to identify 236,507 sessions (= 83.2 %) as sessions of users who are native speakers of one of the languages offered on the website at that time (L1 users). 41,297 sessions (16.8 %) were L2 (non-native speakers) user sessions. There were thus significantly more L1 users than L2 users accessing the websites ($p < 0.001$ in a χ^2-test). In addition, this proportion was 50% higher than the corresponding proportion among all Internet users. In 2001, the proportion of native speakers of the site's four languages among all Internet users was 55.9% (www.glreach.com).

Among the 135 valid answer sets of the questionnaire-based study, 113 (= 83.7%) were answers from L1 users, and 22 (=16.3%) from L2 users ($p < 0.001$ in a χ^2-test). The ratio between native speakers and non-native speakers was therefore almost identical in both studies: significantly more L1 users than L2 users accessed the website. It should be noted that during the data collection of the follow-up study, French was offered as an additional language, which should have led to a higher percentage of native speakers compared to the first study. In 2004, the proportion of native speakers of the site's five languages among all Internet users was 58.4% (www.glreach.com).



* The lower line represents the languages that were "L1 languages" (English, German, Spanish, Portuguese) in the logfile analysis, data from 2001; the higher line represent the languages that were "L1 languages" (idem + French) in the survey, data from 2004

Figure 19. Proportion of L1 and L2 Users who Accessed the Website/Participated in the Survey

These findings seem to clearly reveal the impact of users' language-based cognitive effort on website access. Yet a limitation to such conclusion is the lack of control for a potential bias of the hyperlink distribution. In fact, for a more precise evaluation of the impact of language, insight into the number of links leading to the website is required. In contrast to culture, language is a much more visible and tangible characteristic of websites, in particular if websites are judged from a first glance. Our analysis of information flow focuses on language as a determinant of information flow.

The wide-spread use of L2 information presentation also gives rise to a number of further questions: Does the users' perception of the amount of information offered in their native languages affect their attitude towards L2 language websites? Do their proficiency levels have an impact on user satisfaction?

This chapter investigates the extent to which language represents a barrier to the flow of information on the Internet. We examine whether mechanisms of demand and supply determine the number of webhosts per language, and to which extent these affect language-related link distribution between websites. Beyond this technical data, this chapter also returns to the question of how language affects the extent to which existing hyperlinks between websites are used. Finally,

we investigate whether or not language-related mechanisms of demand and supply also affect users' attitudes and satisfaction with a website.

Insight from this chapter provides further knowledge about the importance of a website's language(s) when attracting users. This type of information is useful for commercial websites but also for websites with the goal of social benefice. It also provides help with promoting health information among large user populations, diminishing the digital divide.

4.2 LITERATURE REVIEW: THE ROLE OF LANGUAGE (AND CULTURE) WITH REGARD TO HYPERLINK SETTING AND FOLLOWING BEHAVIOUR

Investigations into hyperlink structures and link-following behaviour represent the two major areas of research when examining the flow of information on the World Wide Web.

Investigations of hyperlink structures range from predominantly social analyses to primarily technology-oriented investigations of the Web's structure. Within the perspective of the latter, Broder and his colleagues (2000) studied the global hyperlink structure of the Internet and developed formal concepts such as bow-tie, core, in-and-out components, tendrils or disconnected components. The metrics of hyperlink networks are predominantly based on distances and the connectivity between websites and derived measures, such as centrality or cluster analyses (Adamic and Adar, 2001; Albert, et al., 1999; Baeza-Yates and Castillo, 2001). Common application areas are search engines and bibliometrics.

Following results from Baeza-Yates and Poblete (2003), hyperlink patterns are not necessarily universal. The authors found evidence of hyperlink structure patterns that are specific to one country, namely Chile. Their argumentation suggests that the developmental stage of the regional Web is the major cause for peculiarities in hyperlink patterns. Hence, the impact of culture here on hyperlink setting behaviour is rather indirect.

Despite these findings, evidence supporting the impact of the social environment on hyperlinks was also found. Knowledge about its impact is important when information about quality is derived, e.g. from an analysis of how many links are set and to which websites (Park, 2003). Authors focussing on social processes analysed these in terms of both the source and the result of hyperlink setting behaviour. The term *social processes* embraces a number of different phenomena. Data is obtained from metrics concerning the degree of linkage between websites and interpreted with regard to social and communicational ties. McPherson and his colleagues have, for example, argued that Confucianism is a major reason for homogeneity among South Korean political groups

and the link setting behaviour on their websites (McPherson, et al., 2001). Such interpretations are grounded in the belief that "hyperlink structures are likely to be designed, sustained, or modified by website creators to reflect their communication choices and agendas" (see also Jackson, 1997; Park and Thelwall, 2003). In many studies the frequency or intensity of hyperlinks are used to measure the salience of a topic within a social community (e.g, Adamic, 1999; Danowski and Edison-Swift, 1985; Rice and Barnett, 1986).

With respect to the *"social"* consequences of hyperlinks, the vast majority of studies look at the generation of trust and loyalty. Palmer, Bailey and Faraj (2000) showed how the number of links affects the trustworthiness of a website. In a similar manner, Terveen and Hill (1998) examined the number of hyperlinks between websites as an indicator of the quality of sites. The authors found that hyperlink connectivity had a significant relationship to experts' quality judgments of sites. Similar results are provided by Park (2002). The authors found that a site's incoming centrality had a significant impact on website access and perceived website credibility. Incoming centrality (or indegree centrality) is "calculated based on the number of hyperlinks a Web site receives from the other sites". It is opposed to outdegree centrality which is "determined with the number of hyperlinks originating from a site" (based on Freeman, 1979; Park, 2003). A comprehensive of social/hyperlink network analyses can be found in Park and Thelwall (2003).

With respect to the impact of language and culture, first studies have been conducted. In Kralisch and Mandl (2005) we proposed a framework that explained how culturally determined social networks, information need, and hierarchical structures determine website centrality, connectivity, and the motivation for link-setting behaviour. For example, cultures where individual opinions are highly accepted and valued are much more likely to set links to websites of opposite-minded content than cultures with strong collectivistic ties (see alsoSunstein, 2001). McPherson et al. (2001) argue that *"Confucianism"* had a large impact on strengthening homogeneity among South-Korean political actors on the Web, which is mirrored in the hyperlink connectivity between their websites.

Bharat et al. (2001) and Halavais (2000), studied the role of geographic borders and language affiliation on link setting behaviour. It was shown that the number of links within a country domain is generally much higher than towards any other country domain. Results revealed strong geographical connections[43], yet they were sometimes overridden by language affiliation (e.g. Brazil –

[43] About 90 percent of the hyperlinks on U.S. websites link to other American sites. In Europe, between 60 and 70 percent of the hyperlinks are directed to other national websites. Among the remaining hyperlinks, 70 percent link from Europe to U.S. Web sites.

Portugal) (Halavais, 2000). Other authors investigated similar aspects (Barnett, 2003; Brunn and Dodge, 2001; Zook, 2001).

Bharat's and Halavais' studies are a first indicator of the potential impact of language: websites in different languages are less connected than websites in the same language. However, the data analysed in these studies was aggregated on the national level and therefore provided limited insight into the role of language.

Besides the characteristics of link setting behaviour, the flow information on the Internet also depends on how the links are used. Pirolli and Card's Theory of Information Foraging (Pirolli and Card, 1995; Pirolli and Card, 1999 - see section 1.2.2.1.1) again provides a very important contribution. The cost-benefit trade-off of following hyperlinks is directly applicable to our research. Based on the Theory of Information Foraging, other authors developed further predictive models in following years. The number of available links, the number of previously accessed pages, the search goal, and numerous other determinants were identified as having an impact on the perceived value of the information gained. A discussion of these studies' results can be found in Bernard (2000). However, language-related aspects have received little attention within studies of Information Foraging so far.

This chapter therefore aims to investigate the impact of language on link setting behaviour and users' link-following behaviours. None of the previous studies have adopted this type of combined approach. Also, in contrast to previous studies, data aggregation puts a stronger emphasis on the language level rather than the national level. An important focus of our work is to examine how hyperlink setting behaviour depends on whether two websites are in the same or different languages. For an appropriate evaluation of hyperlink setting behaviour we furthermore take the number of webhosts per language into account. The subsequent analysis of how these links are followed represents an extension of results obtained from the studies presented in chapter 2. In addition, behavioural insight is complemented by examining attitudinal variables of link following behaviour and website access.

4.3 EMPIRICAL WORK

4.3.1 Study 8: Behavioural Facts about the Impact of Language on Link Setting and Link Following Behaviour in the Context of the World Wide Web

4.3.1.1 Conceptual Framework and Hypotheses

4.3.1.1.1 Link Setting Behaviour and the Role of Language

The impact of language on hyperlink patterns was investigated from two perspectives. First, the relationship between the number of (native) speakers and the number of webhosts was investigated. We subsequently examined the impact of the number of webhosts in a certain language on the number of hyperlinks coming from websites in that language.

The relationship between the number of (potential) Internet users speaking a certain language and the number of webhosts in that language can be founded on two arguments. The first line of reasoning is based on a market perspective where the number of webhosts in a language follows the number of potential customers/visitors. Due to simple mechanisms of supply and demand, a higher number of potential customers usually attracts a higher number of suppliers. Consequently, the number of webhosts in each language increases with the growing number of users speaking that language as a native or non-native language.

The second argument is the fact that a higher number of (native) speakers increases the number of people who are able to create a website in that particular language. As a result, the number of websites per language should be higher for languages with many speakers than for those with few speakers.

However, the relationship between the number of speakers and the number of webhosts is not necessarily straightforward. Despite a lack of empirical research, it can be expected that the number of users and the number of websites are not directly proportional, due to scale, network, and threshold effects. Scale effects predict that the costs per produced unit decrease with each additional unit. The fact that the value of a network increases more than by n (namely by n^2-n according to Metcalfe's law) with n additional members is called a network effect (Metcalfe, 1995; see also

Odlyzko and Tilly, 2005). In addition, different wealth and heterogeneous education levels as well as discriminatory marketing goals[44] (Grin, 1994) represent further influencing factors.

The connection between the number of webhosts and the number of (potential) hyperlinks is derived from network effects: a larger number of nodes permits more edges between them. Each additional node increases the number of potential edges by n-1 (n= number of nodes) [45]. A higher number of existing webhosts in a certain language consequently leads to a higher number of potential links to and from websites in that language[46]. The number of existing webhosts per language is therefore taken into account in order to allow for the analysis of the impact of language independently from network effects.

We derive the following hypotheses:

H30: The number of in-links from a website that offers information in language y, relative to the number of webhosts in language y, is higher than the number of in-links from a website that offers information in language x relative to the number of webhosts in language x, if language y, in contrast to language x, is one of the languages offered on the website.

The hypothesis is expressed by the following expression:

$$\frac{il_X}{h_X} < \frac{il_Y}{h_Y}$$

il = number of in-links

h = number of webhosts on the Internet

x = a language that is not offered on the target website

y = a language that is offered on the target website

The term "target website" refers to the website whose connectivity with other websites is investigated. It should be noted that this argumentation is based on the assumption that the ratio of the number of in-links per language is distributed equally between all websites, unless influenced by the language of the target website. In the future, more sophistated approaches might take further

---

[44] Even if all speakers of a small language market are bilingual (i.e. they are proficient in the language of a bigger language market), the use of their native tongue can represent an additional service that discriminates a product from others by enhancing its value/perception.

[45] Under the assumption that two websites are linked through only one hyperlink, without considering directionality.

[46] It should be noted that the number of in-links on a particular website increases only by 1 with each additional node.

variables into account. If data is available, the number of website visitors on source websites could be used as a weighting measure. Also, PageRanks – a more accessible type of information - could be applied as a weighting measure (Brin and Page, 1998).

4.3.1.1.2 Link Following Behaviour and the Role of Language

The Theory of Information Foraging predicts link following behaviour based on a trade-off between the perceived costs and benefits of following that link. An examination of the impact of language on link following behaviour consequently requires an analysis of the potential additional navigation costs and values for native and non-native speakers of the language.

There are two major cost aspects that are related to the user's proficiency level in a certain language: cognitive effort and time invested towards understanding (and accessing) the website. Following the Revised-Hierarchy-Model (Dufour and Kroll, 1995) and research results from the field of psycholinguistics (Hahne, 2001), the cognitive effort and time invested with lower language proficiency increase with lower language proficiency (see section 1.2.2.1.1). As a result, users who are not native speakers of one of the website's languages have higher costs for accessing and understanding the information or service offered on that website. This argumentation has already been presented in chapter 2.

In chapter 1 we also mentioned that attitudinal values transmitted through language increase the value of a native language website. Nevertheless, a discussion of whether or not the use of a user's native tongue also increases the website's value, or whether it exclusively diminishes the users' perceived costs, is beyond the purpose of this thesis and would not provide any further insight. In any case, since the use of a native language only rarely diminishes the value of a website (that would be for example the case for negative language-associated values – (Dmoch, 1997; Grin, 1994), it can be argued that information offered in the user's mother tongue always enhances the website's net value: either because it decreases the perceived (cognitive) cost and/or because it adds to the website's perceived value. Therefore, following the Information Foraging Theory, native speakers of the target website's languages are more likely to be able to effectively access the website.

Taking into account the aforementioned network effects, the following hypothesis can be inferred:

H31: The number of website visitors of native language y relative to the number of in-links from webpages in language y, relative to the total number of Internet users with native language y, is higher than the number of website visitors with native language x relative to the number of in-links

linking from websites in language x, relative to the total number of Internet users with native language x, if y is a language offered on the target website.

The hypothesis is expressed by the following expression:

$$\frac{u_X}{il_x * tu_X} < \frac{u_Y}{il_Y * tu_Y}$$

u = number of users/website visitors
il = number of in-links
tu = total number of Internet users
x = a language that is not offered on the target website
y = a language that is offered on the target website

4.3.1.1.3 Reciprocity of Language-related Link Setting and Link Following Behaviour

Language-related link setting and link following behaviour are also characterized by their potential interdependency. Link setting behaviour can be understood as an anticipation of link following behaviour. Resulting from our assumptions about the higher probability of users following links to websites in their native language, it can be assumed that hyperlinks are much more often set to link two websites of the same language than to link websites of different languages. It was also argued that link following behaviour is furthermore affected by the number of existing links. If link setting behaviour anticipates link following behaviour and if link following behaviour is affected by the number of existing links, link setting and link following behaviour mutually reinforce each other: a higher number of users per language leads to a high number of links attracting more users; a lower number of users leads to a lower number of links, preventing additional users from joining (this part of) the Internet. Due to this reciprocal impact, the effect of the number of Internet users per language is expected to be disproportionate, i.e. not of linear character.

Secondly, as a result of a lower number of direct links leading to websites in other languages, it is likely that non-native speakers will have to follow more links to access the target website. In accordance with the Information Foraging Theory, this increases the costs for non-native speakers, decreasing their likelihood of using that website.

Figure 20 illustrates the role of language as a barrier to information, with regard to its impact on the number of hyperlinks and on the number of website visitors.

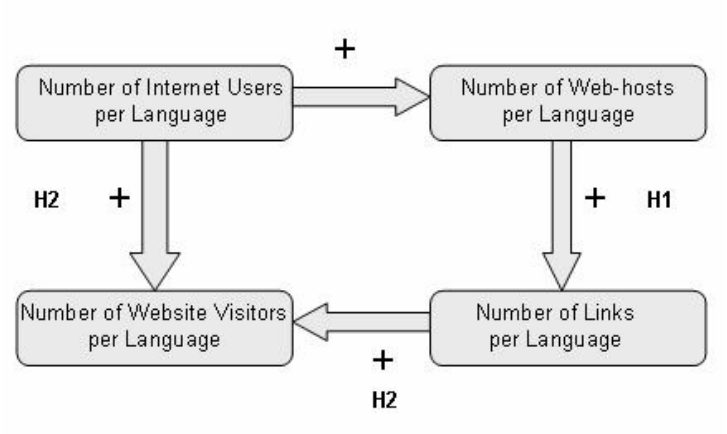


Figure 20. The Role of Language as a Barrier to Information Access on the Internet

4.3.1.2 Method

4.3.1.2.1 Materials and Apparatus

Our study is based on data from website A. Data was obtained from the website's logfile as well as from a web crawler.

The web crawler is based on Jobo (www.matuschek.net) and was developed at the University of Hildesheim[47]. The crawler queries search engines to collect information about other websites and their links to the website investigated. For each of these links, the dataset contains the URL of both the source and target page and their language. For this analysis, all webpages are considered independent objects regardless of potential relationships.

A language identifier was integrated into the crawler for its obvious purpose. We chose Ngramj (http://sourceforge.net/projects/ngramj/), which is based on an algorithm using n-grams of characters (Cavnar and Trenkle, 1994). In the cases where pages contain text in more than one language, it is assumed that there is one main language and the results from the system are used (see also Martins and Marió, 2005).[48]

Due to uncertainties involved in automatic language identification, we also cross-validated the links found by the crawler with the external referrers resulting from the analysis of the responding web-site's logfile. We analysed logfile data from the months of February, March, and April 2005. It should be noted that distinct external referrers only contain in-links that have been used at least

[47] Credits go to Dr. Thomas Mandl.

[48] For confidentiality reasons we are not able to provide examples of the data we obtained from the web crawler.

once. The assumption of having a representative collection of in-links on the server log is more and more justified the higher the number of website sessions. For the purpose of this study more than 90,000 sessions a month were analysed. External referrers that indicated the use of a search engine were excluded.

Information about the website's visitors and their native languages is inferred from the website's logfiles, following the procedures described in section 1.3.2.1.1. Data about the number of hosts and Internet users per language are obtained from public statistics (Languages and Internet UNESCO Culture Sector: http://portal.unesco.org/culture/en/ev.php-URL_ID=21296&URL_DO=DO_TOPIC&URL_SECTION=201.html and www.glreach.com). If not indicated otherwise, these data are from 2005.

4.3.1.2.2 Design

In order to determine the impact of language as a barrier to information flow, data about L1 and L2 websites are evaluated with regard to the number of website visitors, webhosts, and in-links. Due to the low number of languages evaluated, we limit our calculations to simple comparisons of L1 websites/users and L2 websites/users. The L1 group consists of English, French, Spanish, German, and Portuguese. As representative languages of the L2 group, Japanese, Chinese, and Russian were chosen. With regard to their number of Internet users, this language sample represents a diversified mix (see Appendix A-4.1). In fact, the first four languages with the highest number of Internet users are two L1 and two L2 languages.

Where user behaviour was analysed we restricted the analysis to sessions and treated a session as equivalent to a user – in line with analyses in chapter 2.

Numbers of websites, hosts, and webpages per language are analysed in a way where the numbers for each language group is considered relativly to the number of the other language groups. Consequently, analyses are based on approximate numbers and ranking orders, instead of absolute numbers. This allows us to treat the numbers of websites, hosts, and pages as the same variable and assume a similar distribution of the number of websites, webhosts, and webpages with regard to one specific language group. Such an assumption is necessary due to the divergi ng availability of data.

4.3.1.2.3 Procedure

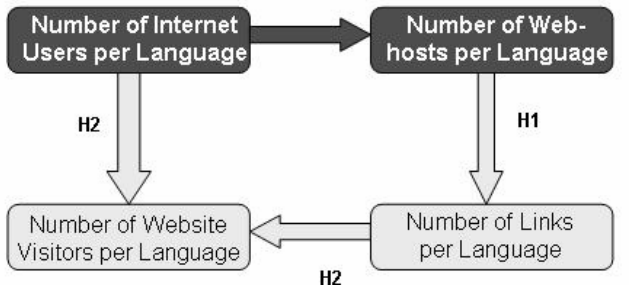
Crawling and log-file analyses were carried out through separated data sets for three months, namely February 2005 through April 2005.

4.3.1.3 Results

The following analyses examine the relationships proposed by our model and hypotheses in a stepwise manner. In each step it is first investigated whether or not a linear relationship exists (*"absolute"*). If this does not apply we then test how the relationship is influenced by the amount of the determining factor (*"relative"*) (e.g. to which extent the ratio webpages/Internet users per language decreases/increases with a growing number of Internet users.) The second step corresponds to the mathematical calculation of the first derivation. Since its result would be zero with a linear relationship, "calculation"/visualization (see below) is only senseful with non-linear relations. Particular attention is paid to the comparison of L1 (offered on the target website) and L2 (not offered on the target website) languages.

In addition to the calculations for testing the proposed two hypotheses, we also present further details in order to provide a more comprehensive picture of the role of language. It should be noted that the nature/linearity of the relationship is only roughly estimated based on (the ranking of) the results. More sophistacted calculations were not considered appropriate due to the very resticted number of languages investigated. In all cases, we added an approximate curve to the results' visualization in order to facilitate the interpretation of the results. However, results are only to scale on the y-axis. Results on the x-axis visualize the ranking order with each language mapped on the x-axis with equal distances from the adjacent languages (according to the ranking order). Exceptions that do not match the regularities found are marked with a circle.

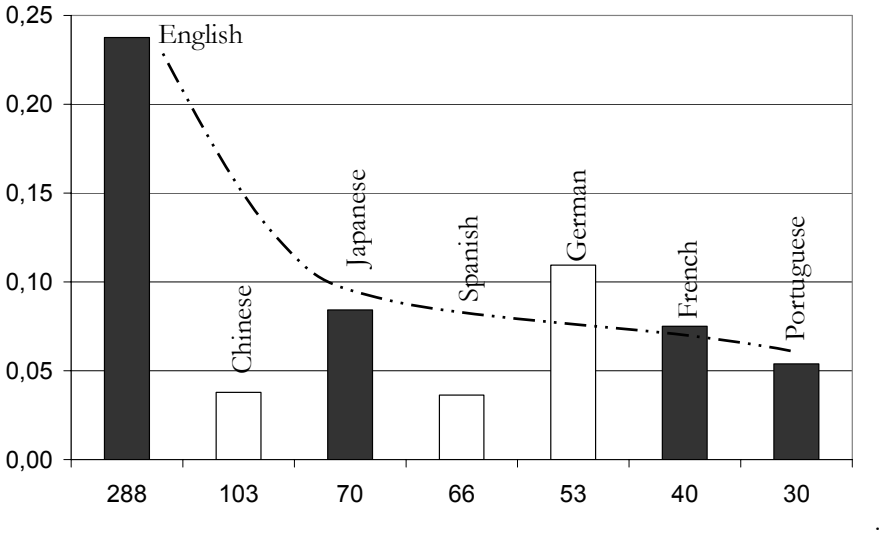
4.3.1.3.1 The Number of Internet Users and Webhosts



This first investigation addresses the question of whether or not the number of webhosts is determined by the number of Internet users. As such it tests for existing mechanisms of supply and demand and provides background information for the subsequent analysis of the number of existing hyperlinks per language. In contrast to the following calculations, figures do not refer to the website investigated, but rather to the entire Internet.

Absolute. Results (Appendix A-4.1) reveal that the number of webhosts per language follows the number of Internet users: the number of webhosts increases with an increasing number of Internet users in the respective language. No fundamental divergences between the L2 languages and the L1 languages were found (see Appendix A-4.1). However, three exceptions were encountered in the analysis: Chinese, Spanish, and German. The nature of their exceptional status is described more in detail below. The relationship between the number of Internet users per language and the number of webhosts is not linear. Subsequently we therefore examined the extent to which the ratio of users and webhosts is determined by the number of Internet users.

Relative. Figure 21 depicts the ratio between the number of webhosts and the number of Internet users per language as a function of the number of Internet users. Webhosts and Internet users are measured as the percentage of the total number of webhosts and Internet users, respectively. A value of 1 would therefore represent a balanced relationship between webhosts and Internet users per language (for exact data see Appendix A-4.1).



X-axis: Number of Internet Users in Millions (not to scale)

Y-axis: Percentage of webhosts/percentage of Internet users

Curve is approximate and indicates the general tendency.

White bars indicate exceptions from the general tendency (see text below)

(Data for 2003; Source: www.glreach.com)

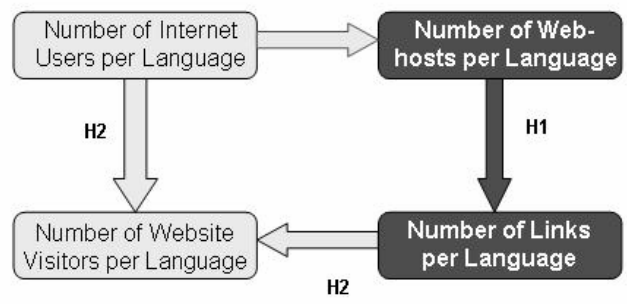
No data was available about the number of Russian webhosts

Figure 21. Number of Webhosts/Internet Users as a Function of the Number of Internet Users

The numbers show a decreasing ratio of webhosts per Internet user for a decreasing number of Internet users. The percentage of English hosts is almost twice as high as the large number of native speaking English Internet users, yet Portuguese hosts represent only a third of the total number of users in that language .

The three exceptions that were found in the first calculations appear here again: these are the Chinese and Spanish groups with an extremely low number of webhosts compared to the number of Internet users, and the German group with a comparatively high number of hosts. In other words, there are fewer Spanish and Chinese webhosts than their number of Internet users would allow us to believe, and more German webhosts than expected.

4.3.1.3.2 The Number of Webhosts and In-links (Hypothesis 30)



This calculation refers to the number of hyperlinks per language that are directed to the websites investigated.

Two datasets were used for this analysis: data from a web crawler and data about distinct referrers from the server-log. The web crawler revealed 4,220 links pointing to pages within the website. The usage analysis revealed between 5,191 (April 2005) and 7,748 (February 2005) distinct links. Whereas language results from the web crawler are based on automatic language identification, top-Level-Domains serve as proxies for the source page language of in-links identified through referrers[49]. The overlap between both sets is only about 20%. This means that the search engine queried by the crawler has not registered all pages linking to the site, and many links known to the search engine were not used.

Due to the low overlap between the two methods, results from both analyses are considered. The number of referrers from a website in a particular language and the number of in-links from web-

[49] It should be noted that such an approach leads to some uncertainty since a .com website is not necessarily an English language website and a .de website does not always offer information in German.

crawling are shown in table 17. For a better comparison we listed the global percentage of websites per language in the last column.

Table 17. Source Page Languages (Feb 05-Apr 05)

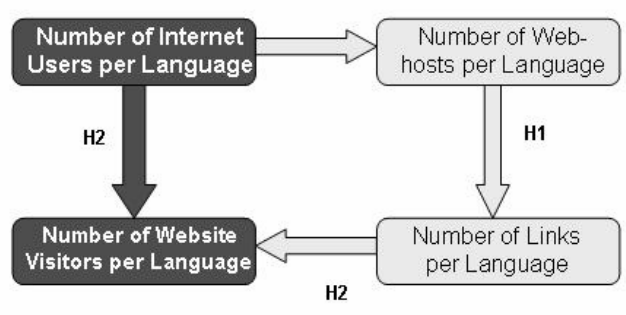
	Referrer Analysis			*Web-Crawling*			*Internet*
Source page language	*Referrers[a)]*	*Percentage of referrers*	*il / h*	*In-links*	*Percentage of in-links*	*il / h*	*% of websites*
English	1,275	16,46%	**0,24**	2,247	53,20%	**0,78**	68%
French	25	0,32%	**0,11**	74	1,80%	**0,60**	3%
German	5,912	76,30%	**13,16**	1,436	34,00%	**5,86**	6%
Spanish	17	0,22%	**0,09**	14	0,30%	**0,13**	2%
Portuguese	67	0,86%	**0,62**	5	0,10%	**0,07**	1%
Japanese	7	0,09%	0,02	0	0,00%	0,00	6%
Russian	26	0,34%	0,18	0	0,00%	0,00	2%
Chinese	7	0,09%	0,02	2	0,00%	0,00	4%

Hypothesis H30 predicts more in-links for L1 languages (English, French, German, Spanish, Portuguese) relative to their global share of webhosts than for L2 languages (Japanese, Russian, Chinese).

Data from web-crawling confirms the hypothesis for all languages. It should be noted, however, that the web crawler did not find a significant number of in-links for each of the L2 languages. According to the referrer analysis, a higher share of L1 users compared to L2 users applies if Russian is not taken into account.

Furthermore, the data reveals that German is the only language where both data sources disclose a higher share of in-links than expected from the number of websites.

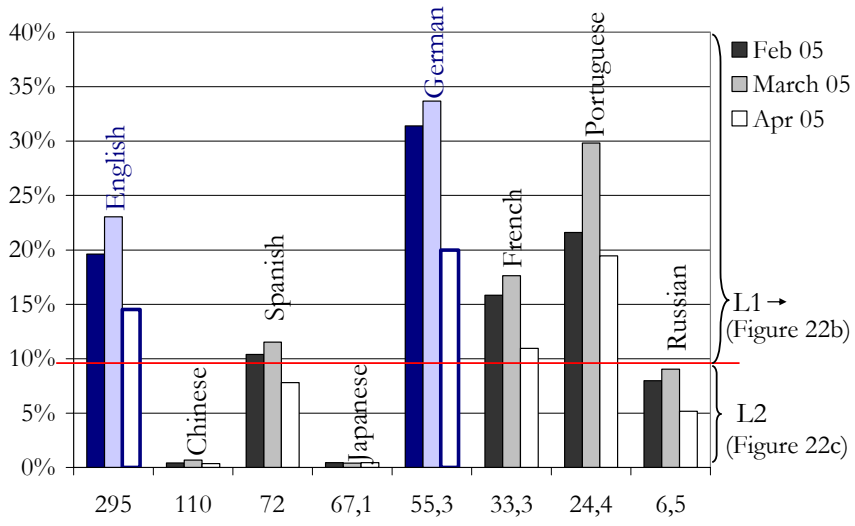
4.3.1.3.3 The Number of Internet Users and Website Visitors (Hypothesis 31: part 1)



This section includes a partial replication of analyses that were conducted in chapter 2. Results obtained from studies presented in chapter 2 and in the introductory part of this chapter revealed a higher proportion of L1 users to L2 users than one would expect from their respective Internet user distributions (For data see Appendix A-4.1).

Absolute. Results from this data set reveal that there is no simple direct relationship between the number of worldwide Internet users and website visitors if all investigated languages are considered. Yet a systematic relationship can be found if the users are separated into groups of non-native speakers and native speakers. Regardless of the number of Internet users per language, in every case L2 website visitors are less represented than any L1 language. These findings confirm the aforementioned outcomes. Within the groups there is again no linear relationship between the number of Internet users and the number of website visitors.

Relative. Figure 22a illustrates the percentage of website visitors per language (relative to the total number of Internet users) as a function of the number of Internet users per language.



X-axis: Number of Internet Users in Millions (not to scale)

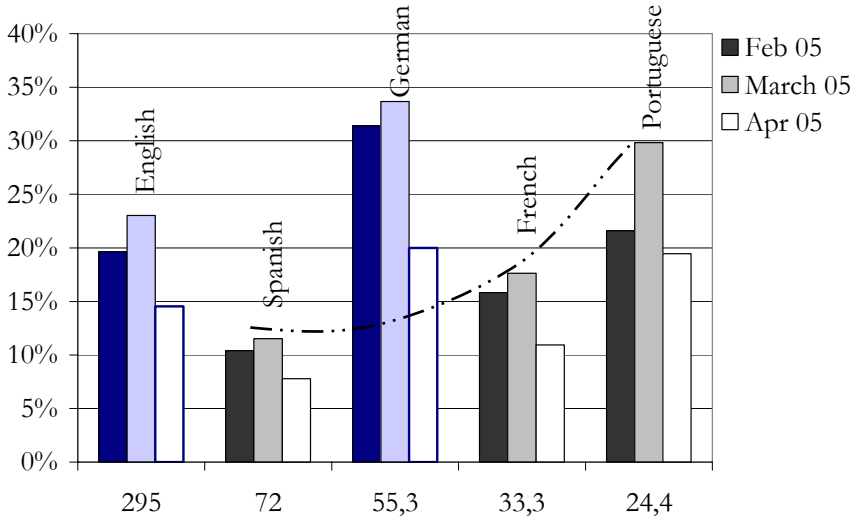
(Data for 2005; Source: www.glreach.com)

Y-axis: Percentage of Website visitors/1000 Internet users

Blue bars indicate exceptions from the general tendency (see text below)

Red line separates L1 languages from L2 languages.

Figure 22a. Percentage of Website Visitors/1000 Internet Users as a Function of the Number of Internet Users (L1 and L2 languages)



X-axis: Number of Internet Users in Millions (not to scale)

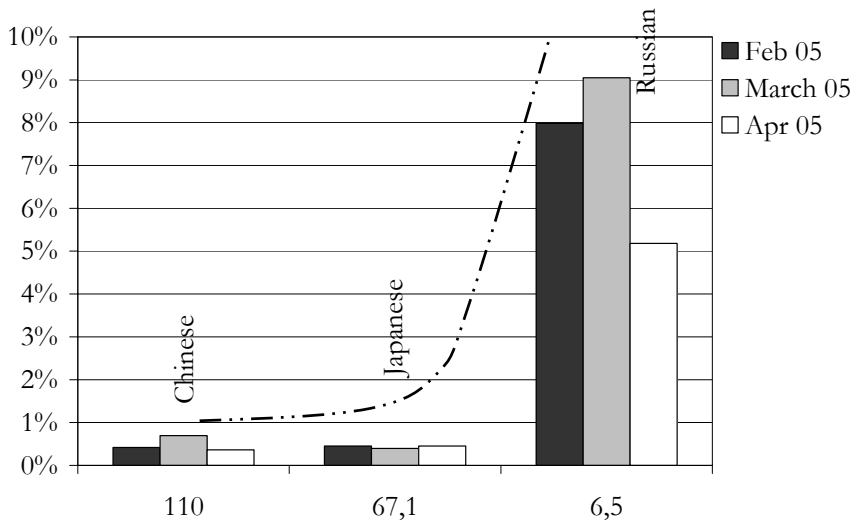
(Data for 2005; Source: www.glreach.com)

Y-axis: Percentage of Website visitors/1000 Internet users

Curve is approximate and indicates the general tendency

Blue bars indicate exceptions from the general tendency (see text below)

Figure 22b. Percentage of Website Visitors/1000 Internet Users as a Function of the Number of Internet Users (L1 languages)



X-axis: Number of Internet Users in Millions (not to scale)

(Data for 2005; Source: www.glreach.com)

Y-axis: Percentage of Website visitors/1000 Internet users

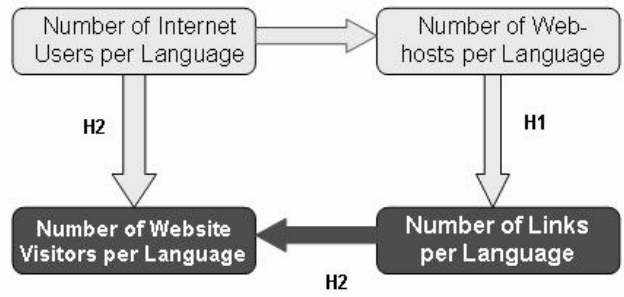
Curve is approximate and indicates the general tendency

Figure 22c. Percentage of Website Visitors/1000 Internet Users as a Function of the Number of Internet Users (L2 Languages)

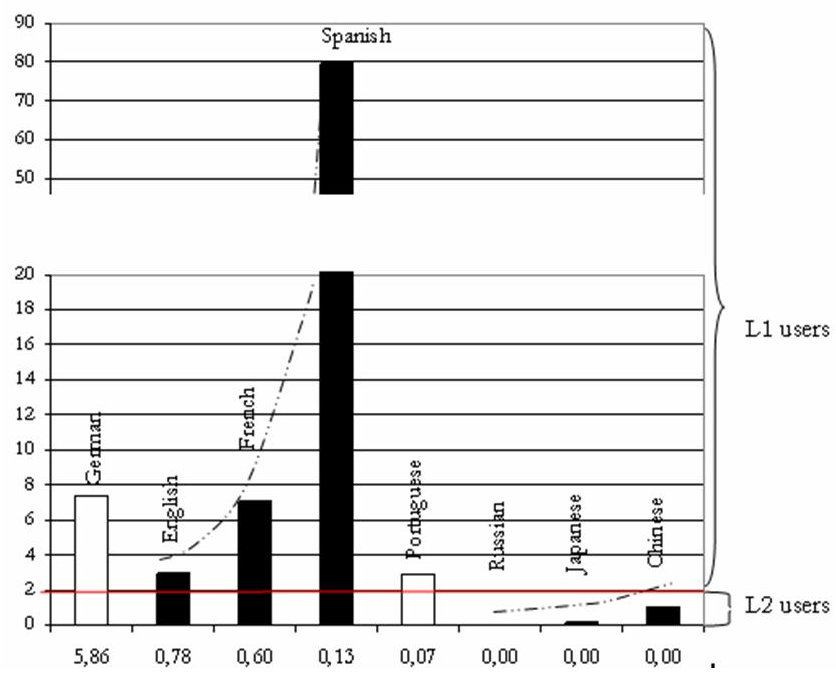
It can be noted that the number of L2 website visitors is without any exception lower than the number of L1 website visitors – despite the fact that the number of Chinese and Japanese Internet users is higher than almost every L1 language group (see Appendix A-4.1).

Interestingly, results for both groups suggest an increasing percentage of website users (relative to the number of Internet users!) with a decreasing number of Internet users. In other words, the lower the number of Internet users, the higher the percentage of users who visit the website. Two major exceptions appear in the native speaker group: the English and the German group are represented by a higher number of website visitors than expected from the respective number of Internet users.

4.3.1.3.4 *The Number of In-links per Language and the Number of Website Visitors (Hypothesis 31: part 1)*



Calculations of the relationship between the number of website visitors per language and the number of in-links are exclusively based on the referrer analysis due to the continuously negligible number of L2 in-links in the results obtained from web-crawling.



X-axis: Number of Website Visitors per Inlink (not to scale)

Y-axis: Percentage of inlinks/percentage of webhosts

Curves are approximate and indicate the general tendencies.

White bars represent exceptions.

Red line separates L1 and L2 languages.

Figure 23. Number of Website Visitors per In-links

Absolute. The numbers show a higher share of website visitors per in-links for L1 users than for L2 users in all cases (see Figure 23).

Relative. Since languages are ordered by their numbers of in-links/webhosts in Figure 23 (from high to low, also see Appendix A-4.1 for detailed numbers), it can also be inferred from the figure that the ratio of website visitors/in-links increases with a lower number of in-links/webhosts when L1 and L2 users are calculated separately. Exceptions are German and Portuguese.

Besides the mere comparison of native speakers and non-native speakers, more detailed examinations of language-related link following behaviour were conducted.

Language of source and target page. Based on referrer analysis and usage analysis, we carried out further analyses regarding the language of the user, the source, and the target page.

As depicted in table 18, referrers with country-specific links are always the most used by users with the corresponding mother tongue (e.g. .jp by Japanese). .de-in-links are the only exception since they were used more by English native speakers than by German native speakers.

Table 18. Which Language Group Uses the Existing In-links (ordered by Top-Level-Domain)?

	Engl.	*Fren.*	*Ger.*	*Span.*	*Port.*	*Jap.*	*Rus.*	*Chin.*	*Other*	*Total*
.com	**86%**	3%	1%	2%	1%	0%	0%	2%	5%	100%
.edu	**96%**	1%	0%	1%	0%	0%	0%	0%	2%	100%
.org	**82%**	5%	3%	1%	2%	0%	0%	1%	8%	100%
.uk	**87%**	2%	0%	1%	2%	0%	0%	0%	5%	100%
.ca	**80%**	10%	0%	0%	5%	0%	0%	0%	5%	100%
.fr	10%	**42%**	29%	6%	0%	0%	0%	6%	6%	100%
.de	**59%**	3%	14%	6%	4%	0%	0%	1%	12%	100%
.mx	14%	0%	0%	**86%**	0%	0%	0%	0%	0%	100%
.es	14%	0%	0%	**86%**	0%	0%	0%	0%	0%	100%
.br	3%	0%	0%	3%	**94%**	0%	0%	0%	0%	100%
.jp	29%	0%	0%	0%	0%	**71%**	0%	0%	0%	100%
.ru	6%	1%	3%	0%	1%	0%	**65%**	0%	23%	100%
.cn	8%	0%	0%	0%	0%	0%	0%	**77%**	15%	100%
.nl	0%	1%	0%	3%	0%	0%	1%	0%	**95%**	100%
.be	5%	1%	0%	2%	1%	0%	0%	0%	**90%**	100%

Data from April 2005

Bold: Language group with the highest percentage per in-link

However, most country-specific referrers have little importance in an overall view, as shown in table 18. The .de in-links are the most used in-link (grey cells in table 4) within all language groups. (exception: .ru for the Russian group).

Table 19. Which In-links are Used within the Language Groups?

Data from April 2005
Bold: In-links with the highest percentage per language group
non-mentioned Top-Level-Domains are used less than 2%

	Engl.	*Fren.*	*Ger.*	*Span.*	*Port.*	*Jap.*	*Rus.*	*Chin.*	*Other*
.com	7%	**5%**	0%	1%	2%	**6%**	1%	**15%**	2%
.edu	**12%**	1%	0%	1%	1%	1%	0%	1%	2%
…..									
.de	**41%**	**42%**	**51%**	**44%**	**40%**	**50%**	**36%**	**60%**	**48%**
.ru	0%	0%	0%	0%	0%	0%	**47%**	0%	0%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%

Due to the outstanding role of English and German, we analysed the link following behaviour of visitors directed to the website through a .com or a .de-in-link in more detail as examples of English and German language referrers. In this context, the question of whether users switch from an English (German) webpage to a webpage in their native language (if this language is offered), is raised.

Table 20. To Which Language Version of the Website Do .com In-links Lead?

Data from April 2005

Bold: Language group with the highest %age per language version

target page→	*German*	*English*	*French*	*Spanish*	*Portuguese*
User↓					
English	0%	**99%**	0%	0%	0%
French	6%	**94%**	0%	0%	0%
German	**56%**	**44%**	0%	0%	0%
Spanish	6%	**42%**	0%	**48%**	3%
Portuguese	0%	**29%**	0%	**12%**	**59%**
Japanese	11%	**89%**	0%	0%	0%
Russian	0%	**100%**	0%	0%	0%
Chinese	2%	**98%**	0%	0%	0%
Other	3%	**97%**	0%	0%	0%

Table 21. To Which Language Version of the Website Do .de In-links Lead?

Data from April 2005

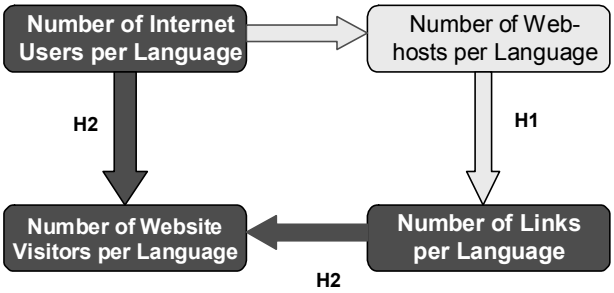
Bold: Language group with the highest %age per language version

target page →	*German*	*English*	*French*	*Spanish*	*Portuguese*
User ↓					
English	3%	**96%**	0%	1%	0%
French	10%	**53%**	**37%**	0%	0%
German	**84%**	**16%**	0%	0%	0%
Spanish	6%	**41%**	0%	**51%**	2%
Portuguese	1%	**37%**	0%	5%	**57%**
Japanese	3%	**92%**	1%	3%	1%
Russian	5%	**95%**	1%	0%	0%
Chinese	2%	**96%**	0%	1%	1%
Other	7%	**91%**	1%	1%	0%

Again outcomes reveal the exceptional role of English. More in depth examinations of the data reveal interesting and important details: divergences in link-following behaviour between L1 users and L2 users appear. In fact, the vast majority of L2 users visit the English version of the website –

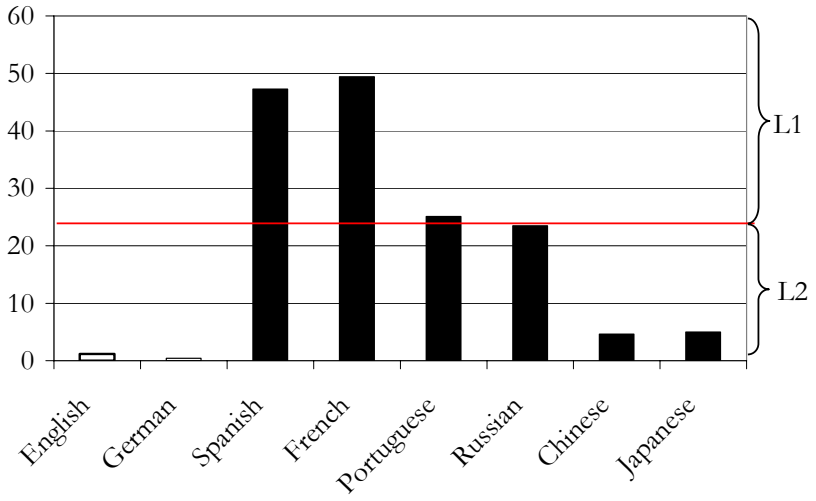
regardless of the language of the source website. In contrast, only about half of the L1 users opt for the English version whereas the other half navigate through their native language website version. Interestingly, the split between English(!) and native languages among L1 users appears to occur independently from the language of the referrer. Rarely do German native speakers use .de in-links for visiting the English version.

4.3.1.3.5 The Number of Website Visitors and the Number of Links per Language and the Number of Internet Users per Language



Hypothesis 31 compares the number of L1 and L2 website visitors relative to the number of Internet users per language and to the number of in-links per language. After having investigated one-way-relationships, the following results take both determinants (Internet users and in-links) into account.

The results first reveal a very low share of website visitors per (webhost * Internet users) for the English and German language group. These are the two languages with a considerably higher number of in-links. Not considering these two languages, all L1 languages have a higher share than the L2 languages. However, the Russian group's share is only slightler lower than that of the Portuguese group (see Figure 24; for data see also Appendix A-4.1). It should be noted that Russian in-links were also over-represented when compared to the number of webhosts (see above).



X-Axis: Languages

Y-Axis: Number of Website Visitors/1000 Internet Users * Percentage of In-links

White bars represent exceptions.

Red line separates L1 and L2 languages.

Figure 24. Number of Website Visitors per Webhost and Internet User

4.3.2 Study 9: Attitudinal Data about the Impact of Language on Website Satisfaction in the context of the World Wide Web

4.3.2.1 Conceptual Framework and Hypotheses

Complementarily to the first study, this study aims to investigate attitudinal aspects. We examine the role of language as a determinant of users' satisfaction with a website. The impact of language is expected to be twofold.

First, a user's language skills affect the perceived cognitive effort required when using the website. This aspect of language's impact has been described and examined at various occasions throughout this thesis. In contrast to previous analyses, users' cognitive effort is measured here in terms of effort saved. *"Saved effort"* relates to the difference in cognitive effort between the use of a user's native a language and the use of a non-native language. In other words, native speakers may differ in the cognitive effort saved dependent on their foreign language skills. This approach is adopted to reflect the increase in cognitive effort L1 users would have if he/she had to visit a website in a different language. A L1 user who is highly proficient in a non-native language saves less effort using a L1 website than a L1 user with less developed non-native language skills. L2 users' *"saved effort"* is always negative and relates to the marginal decrease in effort if they switched to a L1

website. The user's language skills are the user's subjective evaluation of their language costs and values.

Evidence of a positive relationship between cognitive effort and satisfaction has been found in consumer research and analyses of users of search or decision support tools (e.g. Bechwati and Xia, 2003; Branting, 2004).

We infer the following hypothesis:

H32: Users' perceived language-related saved effort is positively related to their satisfaction with the website.

Along with users' linguistic knowledge, the perceived overall offering in a particular language on the Internet is also expected to have an impact on satisfaction. In their simplest form, economic theories predict that a product's value decreases with an increase in the amount of it offered on the market (see e.g. 1994; Grin, 1996). Transferring this assertion to language issues and the Internet market leads us to expect that the language-related value of a webpage decreases as more pages are offered in that language on the Internet[50]. As a further consequence, users with a large offer in their native tongue are more likely to perceive a lower value in a non-native language website than users with a limited offer in their native tongue. This leads to the following hypotheses:

H33: Users' perceived amount of native-language information in on the Internet is negatively related to their satisfaction with the website.

The suggested impact of language skills and language offers on website satisfaction exhibits strong similarities with Davis' TAM model (Davis, 1993). The "*perceived amount of native-language information"* on the Internet can be thought of as its perceived *"usefulness"*, and the perceived effort saved can be thought of as the opposite of the website's *"ease of use"*. As already mentioned repeatedly throughout this thesis, *"usefulness"* and *"ease of use"* have been shown by Davis (1993) to be significant determinants of the acceptance of technology, in terms of attitudes toward the system and subsequent system use.

---

[50] This approach represents a simplified view on a language's value on the Internet. Other aspects such as link distribution/network externalities are likely to also have an impact. However, we consider these factors to have a smaller impact on users' perceptions than the size of the offering in a given language

According to Davis' results there is furthermore a positive impact of *"ease of use"* on *"usefulness"*. The relationship between *"saved effort"* and the *"perceived Internet offer"* is however more complicated. First of all, *"saved effort"* is a construct that expresses *"ease of use"* from a comparative point of view, in terms of opportunity costs: how much effort is saved by using a particular feature instead of another feature. Applied to the language problem the question refers to how much effort is saved/additionally necessary if the website offers information in an L1/L2 language. If additional effort is necessary, i.e. if the website offers information in a non-native language the *"saved effort"* is negative. The effort of using the native language is considered to be zero. The degree of *"saved effort"* depends therefore only on the non-native language skills. In the case of L1 language information presentation, L2 language skills are represented by English language skills. English native speakers were therefore excluded from our analysis. English was chosen because it can be assumed that English is the most frequent L2 website language for Internet users.

The correlation between *"saved effort"* and *"usefulness"* is expected to be negative. The relationship between *"saved effort"* and *"Internet offer"* is expected to be more complicated: it is not clear at this point whether the correlation between *"saved effort"* and *"Internet offer"* should be positive or negative. An argumentation that emphasizes the fact that the *perception* of the Internet offer might *diverge* from the actual offer leads to a negative correlation between the two constructs. It is based on the assumption that users with a higher *"saved effort"* also have a higher need for websites in their mother tongue. Consequently, their perception evaluates the existing language offered as being insufficient. Thus, the higher the users' perceived *"saved effort"*, the lower the *"perceived amount of native-language information"* on the Internet. Another argumentation focuses on market processes that were described in section 1.2.2.2.1, and applies in those cases to where the users' perceptions correspond with the Internet statistics. Market processes predict that users' language skills develop as a result of the language offered, or that the amount of native languages offered takes the speakers' language skills into account. In the first case, users with few websites in their native tongue can be expected to have better foreign language skills. In the second case, it can be assumed that there are more native language websites provided if the users aren't as proficient in foreign languages. In both cases the *"saved effort"* would be positively correlated with the *"Internet language offer"*.

4.3.2.2 Method

4.3.2.2.1 Participants

Data was obtained from two surveys. We obtained a total of 103 valid answer sets from both surveys. The ratio of native speakers and non-native speakers who participated in survey 1 confirms the results from our previous studies. Among the 35 valid cases 30 were L1 users (85.7%) and 5 L2 users (14.3%). Due to the low number of L2 participants, data analyses were based only on the L1 user group.

The opposite situation was encountered in survey 2. Among the 68 valid cases 12 users were native speakers (17.6%) and 56 (82.4%) non-native speakers. It should be noted that only physicians were allowed to register for the website. Thus, patients did not participate in this survey, which explains – in accordance with our previous findings – the high percentage of non-native speakers. For reasons of consistency, native speakers were excluded from subsequent analyses. Website B offers two main services: a discussion forum and a private consultation service. Since not all users have used both services, analyses of survey 2 were split into two parts.

4.3.2.2.2 Materials and Apparatus

Survey 1 was conducted on website A between April and September 2005; survey 2 conducted on website B between July and September 2005. On website A the questionnaire was announced only on the website, and on website B all registered users received an invitation to participate in the survey via E-mail. The questionnaire on website A was offered in the website's five languages (German, English, French, Spanish, Portuguese); the questionnaire on website B was presented in English and German.

4.3.2.2.3 Design: Measures

Questionnaires for both surveys were based on similar item questions but had to be adapted to the characteristics and features of each website. Details about the questionnaires are provided in the Appendix (A-4.2 and A-4.3).

Saved Effort (SAVEFF). By means of a 7-point Likert scale (6-point in survey 2), users were asked in four questions whether they would lose/save time while surfing and seeking for information, whether they would surf more/less, and whether they would perceive more/less effort if they had to surf on a website in a non-native or native language (dependent on whether they were L2 users or L1 users). Answers ranged from *"I strongly agree"* to *"I strongly disagree"*.

Perceived Amount of Native Language Information Online (ILO). The respective three items contained questions about the amount of general and medical information on the Internet in the user's native language, as well as his/her tendency to communicate in English as a result of insufficient native language information. Again, 7 (6)-point Likert scales were used and answers ranged from *"I strongly agree"* to *"I strongly disagree"*.

Satisfaction (SAT). The survey on website A asked users to evaluate their website by means of a 7-point semantic differential scale that used the following adjective pairs: satisfied/unsatisfied; content/discontent; positive/negative; pleased/unpleased. The 6-point semantic differential on website B was based on Davis' items (1993), used for assessing satisfaction (see Appendix A-4.3).

Questions were recoded where necessary, so that a low number stands for low effort saved (SAVEFF), a perceived low amount of native language information (ILO), and a low level of satisfaction (SAT). Accordingly, a high number expresses the opposite.

The questionnaires assessed further language data such as the users' native languages and their English language skills. With regard to the last question, survey 2 distinguished between active and passive language knowledge .

Data about the number of webhosts was obtained from the same Internet statistics that were employed in the previous study.

4.3.2.3 Results

4.3.2.3.1 Study 9a

Answers concerning the three constructs were tested for reliability, and were deemed sufficient in each case (Saved effort (SAVEFF): Cronbach's α=0.855; Satisfaction (SAT): Cronbach's α=0.875; Amount of native language online information (ILO): Cronbach's α=0.736[51]). Factor Analysis (Principal component analysis, Varimax rotation) was performed on the items composing the three constructs. In line with our expectation the three factors were revealed with the respective items and factor loadings (see Appendix A-4.4).

[51] Question 2 was not considered given the results from factor analysis and reliability tests.

Given the sufficient reliability coefficient and the constraints due to the limited data set, each construct was represented by the mean of the items composing it. The Kolmogorov-Smirnov-Test revealed a normal distribution for each construct (see Appendix A-4.4).

We furthermore tested the appropriateness of the *"saved effort"* construct (SAVEFF) and examined the correlation between the construct and the users' foreign language skills. We chose the language most offered online - English - as an example of a foreign language. The correlation is (almost) high with a coefficient of $r=0.698$ and a level of significance of $p<0.001$. The construct *"saved effort"* (SAVEFF) seems to be an appropriate reflection of a user's perceived linguistic effort.

Given the restricted number of cases for each analysis, we limited the analysis to essential statistical analyses. Our hypotheses were first tested through regression analysis. According to these hypotheses the following equation was derived: $SAT = \beta_1 * SAVEFF + \beta_2 * ILO + e_1$. The two independent variables showed no significant impact on the dependent variable. (See Appendix A-4.4 for detailed results.)

Due to the simplified construct calculation we furthermore tested for bivariate correlations (Spearman) among the single items and between the items and the constructs. Significant correlations between the factors loading on the same factor confirm the results obtained from the reliability tests and factor analysis, and therefore the validity of the chosen items. However, *consistent* significant correlations between items of different factors were not revealed; instead correlations between single items provided indices for potential relationships.

Saved Effort – Satisfaction (H28). No correlation between *"saved effort"* (SAVEFF) and *"satisfaction"* (SAT) with the website was found.

Amount of native language information – Satisfaction (H29). The question regarding how accustomed a user is to find information in the native language is negatively correlated with his/her satisfaction (SAT) with the website ($r= - 0.419^*$, $p=0.27$). This means that a user who is used to visiting L2 language websites is more satisfied with a L1 language website than a user who is used to obtaining information in his/her native language.

Saved Effort – Amount of native language information. We found a significant correlation between the amount of effort saved (SAVEFF) and the perceived amount of medical information (ILO) ($r=0.423^*$, $p=0.35$): The more effort saved (SAVEFF), the lower the perceived amount of

medical information (ILO). This outcome supports the assumption of a language skill related perception of the amount of non-native language offers.

Given the low number of native speakers per language, we refrained from carrying out statistical calculations concerning the divergences between the actual online offer per language and the perceived amount (ILO). Nevertheless, a comparison of the ranking order yields important insight (for detail results, see Appendix A-4.4):

Table 22. Comparison of Perceived Amount of Native Language Information and Data from Internet Statistics (Results from Survey A)

	Results from Survey A *Ranking of Perceived Amount of Native Language Information*	*Internet statistics (UNESCO[52]): Ranking of Number Webhosts per Language*
High	English avg.: 5.89 Std.: 1.01	English percentage of webhosts: 68.4%
	German avg.: 5.56 Std.: 1.28	German percentage of webhosts: 5.8%
	Portuguese avg.: 3.67 Std.: 1.58	French percentage of webhosts: 3.0%
	French avg.: 3.0	Spanish percentage of webhosts: 2.4%
Low	Spanish avg.: 2.58 Std. 1.62	Portuguese percentage of webhosts: 1.4%

With the exception of Portuguese, users' amount of information available online in their native language (ILO) resembles that of the Internet. Interestingly enough the German native speakers' perceptions resemble those of the English native speakers, although the percentage of German language webhosts is considerably lower than the percentage of English language webhosts and close to those of the other languages. It raises the question of an existing threshold in user perception, above which additional native language information online hardly increases the perceived quantitative value.

Since the Portuguese group exhibited divergences between perception and actual language offer, we analysed potential correlations between the users' language skills and their perceptions as formulated above. In fact, results reveal that the Portuguese user group is the group with the

[52] Languages and Internet UNESCO Culture Sector: http://portal.unesco.org/culture/en/ev.php-URL_ID=21296&URL_DO=DO_ TOPIC&URL_SECTION=201.html

highest proficiency level in English (excluding native speakers – avg.:2.80 ; for details see Appendix A-4.4). This finding is in line with the argumentation that the *perception* of the amount of native language information is positively correlated with language skills.

4.3.2.3.2 Study 9b – Part 1

54 non-native speakers used the Discussion Forum on the website and answered the respective questions. The reliability coefficient was deemed sufficient for the constructs of *"saved effort"* (SAVEFF) (Cronbach's α= 0.892) and *"satisfaction"* (SAT) (Cronbach's α=0.950), but turned out to be low for the construct of *"perceived amount of information offer"* (ILO) (Cronbach's α=0.460)[53]. Given this outcome we limited the calculation of constructs based on means of *"saved effort"* (SAVEFF) and *"satisfaction"* (SAT). However, the factor analysis (principal component analysis, varimax rotation) also revealed three intended factors with the respective items and factor loadings here (see Appendix A-4.4). Normal distribution is only given for the construct of *"saved effort"* (see Appendix A-4.4).

In line with the results from survey 1, the amount of *"saved effort"* (SAVEFF) significantly correlates with the users' English proficiency levels. It should be noted that the correlation is stronger with the users' active knowledge of English (r=0.388*, p=0.11) than with their passive knowledge (r=0.322*, p=0.40).

Saved Effort – Satisfaction (H28). In contrast to survey 1, results revealed a significant positive correlation between the two constructs (r=0.313*, p=0.25), corroborated by significant correlations between the single items (see Appendix A-4.4). This correlation confirms hypothesis 32.

Amount of Native Language Information – Satisfaction (H29). Again, in contrast to survey 1, no significant correlation between the items that describe the amount of native language information (ILO) and satisfaction (SAT) was found. Yet, in two cases the perceived amount of medical websites (ILO) correlates negatively with satisfaction items (SAT) (r= -0.257 and r= -0.231) at a significance level of p=0.062 (p=0.090 respectively). This outcome is in line with our hypothesis.

Saved Effort – Amount of Native Language Information. Results revealed no significant correlations (see Appendix A-4.4).

---

[53] The value results from item 2 and 3. Cronbach's alpha would be lower if item 1 was incorporated.

We tested the extent to which participants' perceptions of the amount of online information in their native language corresponds with Internet statistics. Only languages for which statistics were available were included in the analysis.

Table 23. Comparison of Perceived Amount of Native Language Information and Data from Internet Statistics (Results from Survey B)

	Results from Survey *Ranking of Perceived amount of Native Language Information*	*Internet statistics (UNESCO[54]): Ranking of Number Webhosts per Language*
High	Russian Avg.: 5.00	
	German avg.: 4.50 Std.: 1.73	German percentage of webhosts: 5.8%
		Chinese percentage of webhosts: 3.9%
	Spanish avg.: 3.75 Std.: 1.26	Spanish percentage of webhosts: 2.4%
		Russian percentage of webhosts: 1.9%
	Italian avg.: 3.71 Std.: 1.98	Italian percentage of webhosts: 1.6%
	Chinese avg.: 3.67 Std. 1.67	
Low	Portuguese avg.: 3.33 Std.: 0.58	Portuguese percentage of webhosts: 1.4%

As illustrated in table 23, there seems to be a tendency towards a general correspondence between the numbers of webhosts per language and the users' perceptions. The comparison of the ranking order between the perceived native language information online and the actual Internet language offer reveals four corresponding cases and two exceptions. The corresponding cases are German, Spanish, Italian, and Portuguese. One exception regards the Chinese language: the perceived amount of Chinese native language websites is ranked considerably lower than the actual amount. It should be noted however that the ranking order of the perceived amount of native language information online Spanish, Italian, and Chinese groups are very close in their perceptions; a ranking-based evaluation therefore has limited validity in this case. The other exception concerns the Russian language. The higher ranking with regard to the perceived amount of language

[54] Languages and Internet UNESCO Culture Sector: http://portal.unesco.org/culture/en/ev.php-URL_ID=21296&URL_DO=DO_ TOPIC&URL_SECTION=201.html

information is here in line with excellent English skills, which, together with the German user group, are the highest among all language groups (avg.: 1.00). Albeit results are in line with those from survey 1, they are of quite limited validity since only one participant was a Russian native speaker.

4.3.2.3.3 Study 9b – Part 2

For this part of survey 2 we obtained 44 valid answer sets from non-native speakers. Similarly to the previous sub-analysis, reliability coefficients are sufficient concerning the *"saved effort"*-construct (SAVEFF) (Cronbach's α=0.947) and the satisfaction-construct (Cronbach's α=0.978). Cronbach's alpha was not acceptable for the items composing the *"perceived amount of Internet language offered"* (ILO). Accordingly, factor analysis revealed only two factors: the *"satisfaction"*-factor (SAT) and the *"saved-effort"*-factor (SAVEFF) (for detailed results see Appendix A-4.5). In contrast to other constructs and items, the construct *"satisfaction"* (SAT) does not apply to the requirements of normal distribution.

The construct of *"saved effort"* (SAVEFF) again exhibits a significant positive correlation that is stronger with regard to active language skills (r=0.388, p=0.11) than with regard to passive language skills (r=0.322, p=0.40).

In contrast to our previous analyses, no correlation between the main constructs was revealed. However, we found a significant negative correlation between the readiness to communicate in English and the effort saved (SAVEFF) (r=0.327*, p=0.040). From this correlation it can be deduced that people with better English skills have a higher tendency to communicate in English.

A comparison of Internet statistics with the users' perceptions of the amount of information offered in their language (through ranking) revealed an overestimation of the information offered from the Russian and the Portuguese groups. For the Russian native speakers, the same argumentation applies as above. With regard to the Portuguese native speaker group, the divergence can not be explained with language skills, since these users rate themselves as less proficient in English than all other language groups (see Appendix A-4.5).

4.4 SUMMARIZING DISCUSSION OF THE ROLE OF LANGUAGE IN THE CONTEXT OF THE WORLD WIDE WEB

4.4.1 Analyses of Behaviour

Results from our previous studies revealed a lower percentage of L2 users and a higher percentage of L1 users accessing a website compared to the distribution of Internet users. Study 8 provides evidence that discrepancies are due to differences in cognitive effort but are also increased by the web's hyperlink structure.

Outcomes from our study confirm the majority of our argumentation and hypotheses. Thus, language has a major impact on who uses a website: websites reach primarily L1 users in the language(s) provided. The following table gives an overview of the results obtained.

Table 24. Impact of Language on Link Setting and Link Following Behaviour

Hypothesis		*Dependent variable*	*Independent variable*	*L1 > L2?*
H30	Link setting behaviour	Number of in-links	Number of webhosts	Yes
H31 (part 1)	Link following behaviour	Number of website visitors	Number of Internet users	Yes
H31 (part 2)	Link following behaviour	Number of website visitors	Number of in-links	Yes
H31 (part 1&2)	Link following behaviour	Number of website visitors	Number of Internet users & number of in-links	Yes (partly)

Link Setting Behaviour. Hypothesis H30 predicted that within the set of pages linking to a site, there are proportionally more pages in the language(s) of that site. For our health site, this hypothesis can be confirmed despite the exception of Russian in one of the data sets. In other words, websites in the same language are more likely to be linked than websites in different languages. It should be noted here that the findings cannot be explained with economic power: basing explanations on the GDP per capita and the number of webhosts, the Japanese sample should exhibit a much higher number of in-links (www.gleach.com/globstats).

Link Following Behaviour. The same divergences between L1 users and L2 users were also found for the number of website visitors. The difference between L1 and L2 users cannot be explained with the number of Internet users. It furthermore raised the question as to whether the higher number of L1 is result of the higher number of in-links that brings them to the website. As mentioned above, there are in fact more links from websites of the same language. The data

suggests that L1 users are overrepresented compared to L2 users even if the number of links (and therefore chances to accesss the website) are taken into account. Nevertheless, results also suggest that such a comparison is only meaningful if the numbers of in-links do not differ considerably. It was shown that the German and the English groups, the two groups with a extremely higher number of in-links, are also noticeably underrepresented. This outcome indicates that there is a threshold in the number of in-links above which the effect of the marginal increase of additional visitors decreases remarkably.

The results providing evidence for the preference of websites in the user's native language are perfectly in line with the Theory of Information Foraging (Pirolli and Card, 1995; Pirolli and Card, 1999) and the Revised-Hierarchy-Model (Dufour and Kroll, 1995).

It was also shown in our analysis – with the exception of three cases – that the number of webhosts per language follows the number of Internet users in that language. Consequently, there are more websites available for Internet users from language groups that are more represented on the Internet than for Internet users from underrepresented language groups. Furthermore, the relationship between Internet users and webhosts is not linear. Results suggest that the number of webhosts per Internet user is higher for languages with a high number of Internet users than for languages with a lower number of Internet users. This additionally increases the benefit for largely represented language groups. This is a typical network effect.

The fact that Spanish and Chinese native speakers are underrepresented whereas the group of German native speakers is over-represented here is likely due to the impact of e-commerce and the users' purchase power (GDP per capita: Spanish 7100 $; Chinese 7200 $ -which is lower than every other investigated group; source: www.glreach.com based on "The Economist" Country Briefings: http://economist.com/countries/ , latest estimated figures). The clear dominance of English is very likely to be caused the wide-spread language knowledge among speakers of other languages.

Nevertheless, exceptions in our results also indicate that the impact of language can be mediated through other factors. In particular, outcomes concerning the Russian group were often close to L1 user groups or constituted an exception. It is possible that cultural affiliation plays a role. Future research is needed here.

Further interesting insight was revealed by the investigation of the users' "language-switching-behaviour". It was found that L2 users generally navigate on the English website version, independently from the language of the referrer webpage linking to the investigated website. In

contrast, half of the L1 users switch to their native language version while the other half opts for the English version – again independently from the referrer language. Only German users rarely switched to English if they were directed to the website by a German source webpage.

Finally, it should also be noted that within the L1 and L2 groups, respectively smaller language groups are more presented than one would expect from the number of Internet users or in-links. This conclusion is inferred from the relative increase of website users with a decreasing number of in-links or Internet users. Network effects are hence slightly mediated but do not overcome the impact of language.

4.4.2 Analyses of Attitudes

Insight from study 9 is less prominent due to the lower number of participants and few significant outcomes. In particular, study 9b allows limited conclusions. Results should be conceived as first indicators of existing relationships that need further research.

It was found that in all three studies the users' perception of *"saved effort"* as described above reflected their self-rated language skills well. Users with lower non-native language skills perceived a higher *"saved effort"* if the website information was presented in their native language.

The positive correlation between the *"saved effort"* and the users' *"satisfaction"* with the website was confirmed in study 9, investigating L2 users. In the same study significant negative correlations between items describing the *"perceived amount of native language information"* and items describing *"satisfaction"* were revealed. This outcome confirmed hypothesis 33, which predicted higher *"satisfaction"* with the website if the amount of native language information is low.

Similarly, results (from study 8) examining native speakers suggest a negative correlation between the *"perceived amount of native language information"* and *"satisfaction"* with the website. In contrast to the L2 user sample, no correlation was revealed for the L1 user sample that would predict an impact on *"saved effort"*.

Outcomes from all three studies suggest an impact from *"saved effort"* on the perception of the amount of non-native language website offerings. Users with low non-native language skills are more likely to underestimate the amount of native language information available whereas it is the opposite case for users who are very proficient in non-native languages.

To sum up, our studies provide evidence for the crucial role of language when information is accessed on the World Wide Web. Despite the role of English as Lingua Franca of the Internet, information presentation in the users' native language seems to be the most decisive factor for attracting website visitors. Two key issues of representativeness were addressed. First, the numbers of L1 and L2 users who access a website may depend on the fraction of L1 and L2 users in the Internet-user population at large. To allow for this potential influence, we considered current data on global Internet usage.[55] The second factor includes the share of in-links from L1 and L2 websites. These issues were addressed in our first study.

From a behavioural perspective, language represents a double barrier for information access on the Internet. First, the size of the language group disadvantages smaller language groups due to the lower number of webhosts and links. As a consequence, less native language information is offered on the Internet and the respective websites are less connected. Second, with regard to accessing the information on a particular website, it was shown that the languages in which the information is presented influence who accesses the website, possibly due to higher cognitive effort required. L2 users are strongly under-represented even if other network effects are taken into account.

Consequently, the size of the language groups and membership in either the L1 or the L2 user groups has an impact on the number of website users per language group. The impact of website-specific characteristics and e-commerce (e.g. low number of Chinese and Spanish webhosts) related issues are likely to be the reasons for the few exceptions that were encountered in our analysis (e.g. dominance of German in-links). The overall cultural impact of the website should be limited since the health-related topics presented (although originally developed for physicians) are rather universal.

From an attitudinal perspective, non-native language skills seem to be the more important determinant of website *"satisfaction"* in the L2 user group, whereas in the L1 user group the perceived amount of native language information prevails. From the positive correlation between *"saved effort"* and *"satisfaction"*, it can be inferred that higher website satisfaction can be obtained in user groups that are not proficient in foreign languages. From a different point of view: non-native language websites are more likely to generate user satisfaction among users who are highly

[55] These data are coarse-grained – more fine-grained data would require either to ask every Internet user for his/her native language or to derive data from statistics on the distribution of native languages in each country (including immigration and other minority language groups) and the extent to which each language group is represented on the Internet. To our knowledge, such data does not exist at the moment.

proficient in the website's language than among those who are less proficient. The negative correlation between the *"perceived amount of native language information"* and website *"satisfaction"* suggests that L1 websites generate higher satisfaction for user groups that are not accustomed to being provided with native language information. A focus on these user groups in a website's language offer would comply with Grin's "discriminatory language market goals" (see 1.2.2.2.1).

5. CHAPTER:

SUMMARIZING DISCUSSIONS

5.1 SUMMARIZING DISCUSSION ON THE IMPACT OF LANGUAGE AND CULTURE ON BEHAVIOURAL AND ATTITUDINAL ASPECTS OF INTERNET USE

5.1.1 Overview of Studies and their Results

Within the previous chapters we presented a number of studies containing several substudies. The outcomes of our investigations often required detailed descriptions. As a synopsis, the following table provides a simplified overview of the studies presented and their results. The following summarizing paragraphs present more explanations.

Table 25. Overview of Studies and Results

Culture was operationalized through Hofstede's and Hall's cultural index scores. Following abbreviations are used:
UA = Uncertainty Avoidance;
LTO = Long-term Orientation;
PD = Power Distance;
IND = Individualism;
MAS = Masculinity;
CONTEXT = Context specificity;
MONOCHRONICITY = comparison of monochronic vs. polychronic cultures.

Investigations of language compared native speakers and non-native speakers of the languages offered on a website.
L1 = native speakers
L2 = non-native speakers

Study 1	
Question	*Impact of culture on navigational behaviour?*
Results	Significant, positive correlations found for: UA - Quantity of information collected LTO - Time spent per page Monochronicity - Linearity of navigation patterns

Study 2	
Question	*Impact of culture on search behaviour?*
Results	Between cultural group comparisons Search engine preferred by High UA (vs. low UA) Low context (vs. high context) Low LTO (vs. high LTO) Alphabetical search preferred by High PD (vs. low PD) Content search preferred by High context (vs. low context) High LTO (vs. low LTO) High PD (vs. low PD) Within cultural group comparisons: Alphabetical search is in all groups the most preferred search option. The search engine is in all groups the least preferred search option.

Study 3	
Question	*Impact of language on search behaviour?*
Results	Significant mediating effect of domain knowledge within the L2 user group Within L2 group comparisons Search engine preferred by physicians (vs. patients) Alphabetical search preferred by physicians (vs. patients) Content search preferred by patients (vs. physiciants) Within L1 group comparison no clear difference between patients and physicians within the L1 user group

Study 4	
Question	*Impact of language on preferences for information categorization?*
Results	Within L1 group comparison
	verbal cues easier to use (vs. visual cues)
	verbal cues more useful (vs. visual cues)
	Within L2 group comparison
	visual cues easier to use (vs. verbal cues)
	visual cues more useful (vs. visual cues)
Study 5	
Question	*Impact of culture on risk perception and risk behaviour?*
Results	no significant impact of culture (stable concept)
	significant impact of quality of health care system, web experience, domain knowledge on the perception of risk and on the perception of information quality
Study 6	
Question	*Impact of culture (basic and broadened concept) on need for website features?*
Results	Higher need for more information
	low IND (vs. high IND)
	low MAS (vs. high MAS)
	low UA (vs. high UA)*
	users from countries with a good health care system (vs. bad health care system)
	Higher need for comprehension support
	patients (vs. physicians)
	Higher need for navigation support
	users from countries with a good health care system (vs. bad health care system)
	* low UA culture have a higher need to access additional information, whereas high UA culture have a higher need to access existing information
Study 7	
Question	*Impact of culture on privacy issues?*
Results	Power Distance and Individualism need both to be considered
	Distinction between privacy-sensitive and low-privacy sensitive data
	Low PD + high IND cultures have higher privacy concerns (vs. high PD + low IND)
	A more positive attitude towards data disclosure leads to a higher willingness to disclose low-privacy sensitive data

Study 8	
Question	*Impact of language on information access on the Internet (link-setting and link-following behaviour)?*
Results	more links between websites of the same language
	links between same language websites are more used than links between websites of different languages
	with a few exceptions, number of webhosts per language follows the number of Internet users

Study 9	
Question	*Impact of language on user satisfaction?*
Results	lower linguistic effort for L2 users correlates with higher satisfaction
	a lower perceived amount of websites in a certain language correlates with higher satisfaction (L1 and L2 users)
	users with a high foreign language proficiency (FLP) level tend to overestimate the number of websites in their native language
	users with a low FLP level tend to underestimate the number of websites in their native language

5.1.2 Behaviour and Attitudes

Throughout this thesis the impact of language and culture were investigated from behavioural as well as attitudinal perspectives. The theoretical basis for the connection between attitudes and behaviour is provided by Ajzen and Fishbein's Theory of Reasoned Action (Ajzen and Fishbein, 1977; Ajzen and Fishbein, 1980), that was cited on several occasions within this thesis. We also mentioned the Technology Acceptance Model by Davis (1993), which is based on Ajzen and Fishbein's Theory. Two important components are the *"perceived usefulness"* and the *"belief about ease of use"*, which are antecedents of the *"attitudes"* towards a system, which in turn affect system *"use"*. An adaptation of Davis' TAM to a cross-cultural background was introduced by Evers and Day (1997). The model incorporates *"culturally specific design preferences"* as external stimuli.

According to these models, users' attitudes towards a system predict the use of the system, i.e. attitudes result in behavioural consequences. Nonetheless, (a) a direct impact of the users' perception about a system's *"ease of use"* on *"system use"* (without the intermediary position of *"attitudes"* and thus without taking into account *"perceived usefulness"*), and (b) a direct impact of the users' *"culturally specific design preferences"* on *"attitudes"* (without the intermediary position of *"ease of use"* and *"usefulness"*) has been suggested by results from Evers and Day (1997).

According to Ajzen and Fishbein, *"attitude"* is the affective evaluation of behaviour or an objective (Ajzen and Fishbein, 1977). It is usually measured through semantic differentials or Likert scales

which were also adopted for the purpose of our studies. Allport (1935) pointed out very early that attitudes are acquired through socialization processes. This is in line with the above mentioned models and with our assumption of the impact of culture and language.

In view of our empirical work we provided theoretical foundations for the impact of language and culture on cognitive processes, in particular on cognitive burden. However, cognitive processes were not measured directly but inferred from observed behaviour. Behaviour[56] was understood in our thesis as users' actions that are visible to an outside observer. We restricted our research to behaviour that was visible by means of logfile analysis or web-crawling. In study 4, behaviour regarded the use of certain categorization cues and was captured by recording the participants' actions within the experiment by means of a computer. Behaviour was usually a dichotomous variable of an action that either occurred or not. Further measures were inferred through aggregation on the session or group level. Frequency of behaviour was furthermore positively linked to preference: for example, a more frequent use of search engines compared to alphabetical search indicated a higher preference for search engines than for alphabetical search.

Depending on the context of the research design and questions asked, a positive attitude (including beliefs about *"usefulness"* and *"ease of use"*) was interpreted as a preference for a specific feature (study 4); a negative rating of a feature pointed out a high need (studies 5 and 6). Questions and scales were adapted to the respective research questions.

As mentioned in the introduction, a complete picture of language and culture with regard to behavioural and attitudinal aspects of Internet use cannot be provided within one thesis. Nevertheless, the findings illustrated in their respective sections and summarized in the following section represent a useful contribution towards understanding this connection.

5.1.3 The Impact of Culture

According to the theoretical discussion preceding the empirical work, we investigated culture in two ways. We distinguished between the basic concept of culture involving stable cultural aspects, and a broadened concept of culture involving dynamic cultural aspects. These two different concepts were chosen as an answer to the problems of causality, inherent to quantitative cultural research

---

[56] *"Behaviour"* needs to be distinguished from Hofstede's terms of *"practices"* and *"rituals"*. "Rituals are collective activities that are technically superfluous in reaching desired ends […]". The term *"practices"* is an umbrella term that subsumes the term *"rituals"* as well as cultural *"symbols"* and *"heroes"* Hofstede, G. (1991): Cultures and Organizations: Software of the Mind, McGraw-Hill, London..

based on Hofstede's paradigm. Within these two concepts a number of aspects were examined. To determine the impact of culture on behaviour, we analysed the influence of stable cultural aspects. With regard to attitudes, dynamic cultural aspects were added to the analysis.

Analyses presented in chapter 2 investigated the impact of culture on behavioural aspects of website use. The impact of culture is theoretically founded on the fact that culture, as the "software of the mind", determines thinking patterns and the cognitive effort required when using specific systems and features. Results from Luna et al. (2003) support the assumption in order to understand the congruity between the characteristics of systems and their users' culturally determined preferences, in terms of cognitive effort.

Results obtained from our empirical investigations show a clear impact of culturally determined thinking patterns on patterns of navigation behaviour and the use of search options. Outcomes were mainly consistent and complementary, so we were able to develop recommendations for structural website design that consider two key designs for two cultural main groups. Due to the predominantly non-reactive manner of data collection, data about users' perceived *"ease of use"*, *"usefulness"*, and *"attitudes"* were not assessed. However, if our assumption of culture as a determinant of cognitive effort is correct, that in turn results in a higher or lower use of specific systems or features, and findings about the culturally determined use of search options and navigation patterns should correlate with the culturally determined *"ease of use"* of these search options and navigation patterns.

Nevertheless, culturally determined differences were only found in between-group analyses. Within-group analyses of preferences for search options revealed, across all groups, a preference for alphabetical searches over content searches, which in turn was preferred over search engines. Similar preferences were also the outcomes of other studies (Chimera and Shneiderman, 1994). Hence, the question of whether culture should be interpreted as a mediator of otherwise universal preferences can be raised. However, it should be noted that in our case preferences are a behavioural response to cognitive effort caused by specific systems and features. It is not clear what the neural bases are and whether a distinction is necessary, namely between characteristics whose cognitive effort is due to universal human abilities and characteristics whose cognitive effort is strongly affected by culturally determined thinking patterns.

Affective components such as attitudes towards content and design features as well as risk perceptions were investigated in chapter 3. Due to the different context (see also 3.1) dynamic cultural aspects were added. Results obtained in this chapter suggest two main findings. First, the

impact of culture on affective components is less salient than the impact on behaviour. Second, the impact of dynamic aspects is stronger than that of stable aspects (see 3.3.1.3, 3.3.2.3, 3.4).

Nevertheless, study 7 provided evidence for the strong impact of stable cultural values on users' attitudes by analysing the impact of Power Distance and Individualism on data disclosure. Interestingly enough, study 7 differs from study 5 and 6 in terms of the correspondence between stable cultural values and dynamic cultural values. In study 5 and 6 dynamic cultural values were basically independent of stable cultural values, whereas the conceptual framework of study 7 portrayed findings from other works where a strong relationship between cultural values and privacy legislation was found. Due to problems of quantification we were however not able to provide empirical evidence of this relationship nor of its impact. Future research is required here.

Study 7 furthermore provided the link between culturally determined *"attitudes"* (towards disclosing personal data) and the users' willingness to disclose this data as a proxy of behavioural responses. Nevertheless, since we did not have the possibility to investigate the users' actual behaviour of disclosing personal data, insight has to be interpreted with caution.

The question of whether a website information service generates low involvement that does not trigger significant cultural differences was also raised. An argument in favour of the low-involvement explanation is the fact that the risk reduction model (study 5) is weak despite its validation in other contexts. Behavioural responses to users' risk perception are rare and are only found with regard to dynamic cultural aspects. An argument that emphasizes the generally low impact of culture (especially stable aspects) on affective components (here: risk perception) is the fact that even in online shopping situations the effect of culture appears to be weak (Jarvenpaa and Tractinsky, 1999; Ko, et al., 2004).

In accordance with the fact that the website investigated offers information services, the Uncertainty Avoidance dimension seems to be the most important stable cultural aspect. In chapter 2 we argued that Uncertainty Avoidance determines the need of information. Support for the culturally determined information need was found in several outcomes. However, the role of Uncertainty Avoidance appears to be rather complex and requires precise distinctions. In fact, outcomes from chapter 3 suggest that access to existing information needs to be distinguished from the desire for additional information. Access to as much existing information as possible is preferred by members of high Uncertainty Avoidance cultures. The results from chapter 3 reporting reasons for visiting the website as well as the results from chapter 2 where high uncertainty avoidant users were found to collect more information than low uncertainty avoidant

users confirm this. The stronger desire for additional information among members of low uncertainty avoidant cultures is supported by website feature ratings that indicate a higher need of information. These behavioural and attitudinal findings are completely in line with Marcus and West Gould's (2000) results. The authors state that websites with restricted information and a restricted number of links create more comfort for high uncertainty avoidant users than websites with unlimited choice. The reason lies in the fact that high Uncertainty Avoidance users tend to want to acquire as much information as possible, an exhausting feat given unlimited options.

To sum up, according to models that link attitude and behaviour, attitude is an antecedent of behavioural outcomes. It was found that stable cultural values have strongly influenced users' behaviours in terms of search option use and navigation patterns. With regard to the affective components investigated, the impact of stable cultural aspects is weak; the impact of dynamic cultural values prevails.

Even though not explicitly assessed, it can be assumed that stable cultural values mainly determine the ease of use, which has a direct impact on system use. Similar results were also found by Evers and Day (1997). Dynamic cultural values were taken into consideration in the investigation of attitudinal aspects since we expected the current socio-economic settings, which are partly expressed in these values, to have a stronger influence. Due to the fact that the impact of dynamic values was more salient than that of stable values in our investigations, it can be assumed that the users' *"attitudes"* were dominated by the perceived *"usefulness"*.

It remains up to future research to provide a more detailed picture of the interrelationship and interaction between attitudes and behaviour. In particular, non-reactive data collection used for analysing behavioural patterns need to be followed up by studies that combine investigations of behaviour and attitudes. The question of causality in cross-cultural research (Which aspects should be taken into account? Which current socio-economic settings are determined by culture? Which are independent of culture? - see also sections 1.1.3.1 and 1.1.3.4) is strongly linked to this problem and therefore needs to be addressed as well. Finally, although mentioned in a number of methodological papers (e.g., Karahanna, et al., 2002), the degree of culture's impact should be much more a point of interest, in particular with regard to cognitive aspects and behavioural responses: What is universal? What is restricted to cultural groups?

5.1.4 The Impact of Language

In the introduction we distinguished between the user perspective and the provider perspective. Since attitudinal outcomes of language use are mostly important with respect to the users, analyses focus on that perspective.

Similar to the analysis of culture, we considered cognitive aspects of language and the respective behavioural responses in chapter 2. Study 4 and study 9 added data about attitudinal aspects of the impact of language.

According to our findings from study 3, language has a strong impact on the use of search options, however only in a low domain knowledge situation. Both the language and domain knowledge variables affect the amount of cognitive effort involved in the use of a certain search option. Since no differences were found between L1 users of different domain knowledge, and since L2 users with high domain knowledge did not differ from L1 users, it can be assumed that only linguistic cognitive effort above a certain threshold affects user behaviour.

A further behavioural consequence of language use is the fact that fewer L2 users accessed websites. Within the L2 user group, users with high domain knowledge prevailed. The dominance of L1 users can still be found even if the lower number of in-links for L2 users is taken into account. This was the outcome of study 8.

The impact of language and domain knowledge on behaviour was supported by results from study 4. In this study, only participants with low domain knowledge were tested. Visual cues were preferred over verbal cues from all groups. Groups differed however, with regard to which visual cue was preferred. Further support for the impact of language is provided by the fact that L1 information presentation leads to a higher *"usefulness"* and *"ease of use"*. However, the link between these two variables and users' attitudes towards certain categorization schemes is not clear. Moreover, results suggest a direct impact of language on *"attitudes"* that in turn result in behavioural responses. Such a direct connection between the external stimulus and *"attitudes"* was also one of the outcomes in (Evers and Day, 1997).

In study 9 we analysed the role of language from a different perspective with regard to attitudinal outcomes, as *"satisfaction"*. Two constructs that are similar to Davis' *"ease of use"* and *"usefulness"* were introduced. We chose the construct of *"saved effort"* which also reflects the cognitive effort L1 users would have if they had to navigate on a L2 website, since it is believed to reflect antecedents of attitudinal outcomes in a more appropriate manner. In fact, a positive correlation between the two

variables was found. *"Usefulness"* was represented by the users' perceptions of native language information online (*"ILO"*) that was negatively[57] correlated with *"satisfaction"*.

Interestingly enough, *"saved effort"* and *"ILO"* are linked in the opposite way as one would expect from Davis' TAM. This type of relationship also makes us assume an opposite direction of the causal chain, which is in the language case from *"saved effort"* to *"ILO"*. In fact, users with high L2 proficiency skills tend to overestimate the number of native language websites offered, whereas those with low L2 proficiency skills tend to underestimate it. *"Satisfaction"* with a website (as a possible antecedent of website use) is consequently determined by other (competing) websites but also by language skills.

Study 8 provided further remarkable outcomes supporting a possible additional link between the attitude towards a language and the behavioural response. It was found that only about half of the L1 users navigate in the L1 website version after being referred to the website, whereas the other half uses the English website version – independently of the referrer language. Since such a behaviour cannot be explained with cognitive effort, the question of whether low trust and/or a negative attitude towards information in the native language information leads to these behavioural responses, is raised.

From the provider's perspective, language plays a role for market behaviour since it affects the borders of the target group. As such it determines how many users can be reached, and as a characteristic of the users, the e-commerce market. Results from study 8 suggest that the number of webhosts mainly follows the number of speakers, and in some cases also e-commerce figures. The rather exponential character of the relationship between webhosts and native speakers is an indicator of typical network and cost saving effects.

To sum up, relationships found between users' linguistically determined cognitive effort and their use of search options and their cognitive effort and satisfaction, usefulness, and ease of use seem to provide evidence of the consistent role of language with regard to behaviour and attitude, as predicted by the models mentioned above (section 5.1.1). However, that our analyses only examined the relationship between the *perceived* cognitive effort and satisfaction should be taken into consideration. Furthermore, while the impact of language (in combination with domain knowledge)

---

[57] The perception of the Internet offer was negatively correlated. The usefulness of a particular website increases the lower the Internet offer. *"Usefulness"* would therefore be positively correlated with *"satisfaction"*.

is rather high, its role with regard to the users' attitudes is less clear and salient as indicated by weaker and less consistent results.

It is up to future research to conduct analyses as an extension of our outcomes, to combine behavioural and attitudinal aspects within one study. In view of the market orientation, analyses were conducted on the language level. Future research investigating the role of language on the individual level and using different samples might provide further insight. Also, outcomes are often based on coarse-grained analyses due to the lack of general language data.

5.1.5 Language and Culture: Commonalities and Interrelationships

The impact of language and culture on behaviour and attitudes with regard to a number of aspects is intertwined, or represents strong similarities. Behavioural outcomes, with regard to both variables, are predicted by cognitive effort. Whereas the cognitive effort of language is determined through cognitive abilities, the cognitive effort of culture is determined by habitual thinking patterns. Domain knowledge was, for example, introduced as a variable that mediates language abilities, but was also used as a cultural aspect that defined a knowledge-determined culture, i.e. a way of thinking.

Since both language and culture are inherent to the individual, an effect can never be clearly assigned to one of the variables, even though an appropriate design can limit the influence of the variable not analysed. In study 4, we found an impact of language on the users' attitudes towards using certain visual categorization schemes. Since the impact of language in terms of cognitive effort is not obvious in this case, the question of whether these differences are due to culturally determined characteristics of language is raised (see section 1.2.1.2).

The difficult separation between culture and language is reinforced through the fact that language data is often derived from country data.

Interestingly enough, it was found that both language and culture had a rather strong impact on behavioural outcomes, and a less clear impact on attitudinal outcomes. There are two possibilities explaining this phenomenon. First, attitudinal responses are usually more difficult to assess than behavioural responses. This is particularly true with regard to cross-cultural and cross-linguistic research that has to deal with particular challenges (1.3.3). Second, the relationship between attitudes and behaviour is more complex in a cross-cultural and cross-linguistic context. In the models mentioned, attitude that predicts behaviour is a result of perceived *"usefulness"* and *"ease of use"*. In our analyses behaviour was only predicted by the cognitive effort respectively constructs

that correspond to *"ease of use"*. The impact appears to be rather strong. As mentioned above, Evers and Day (1997) also found evidence for such a direct impact in a cross-cultural context.

Attitudinal outcomes seem to be much more influenced by the perceived *"usefulness"*, or directly affected by culturally and linguistically determined preferences. The stronger impact of *"usefulness"* is in line with Davis' TAM model. The direct impact was here again also found by Evers and Day (1997). The direct impact can be explained with the attitudes towards cultures and languages that are acquired over a lifetime, which are often independent of any aspects of *"usefulness"* or cognitive effort, having hardly any rational explanation (e.g., 1.2.2.1.2). An example is the so-called *"country-of-origin effects"* (Dmoch, 1997; Gürhan-Canli and Maheswaran, 2000). It remains up to future research to answer this question.

5.2 LIMITATIONS OF OUR RESEARCH

Limitations of our research can be divided into two main groups: those that are inherent to the methods applied and those that are inherent to the topic investigated.

In our analyses we employed three main techniques: (1) logfile analysis and web crawling, (2) online surveys and (3) laboratory studies.

Logfile Analyses. Limitations of logfile analysis regard, in particular, session reconstruction and the lack of control of personal and situation variables. This concerns our studies for several reasons.

Firstly, logfile analysis was used to investigate navigation behaviour in general and search behaviour in particular. Due to the lack of insight about the user's search goal (see section 2.2.2), we had to base our analyses on the assumption that the users' search goal is equally distributed across cultural and linguistic groups. Nevertheless, it is well-known that the choice of navigational tools and the strategies in which these tools are employed are related to a user's search goal (e.g., Allinson and Hammond, 1989; Shneiderman, 1997). For example, it is generally acknowledged that known-item searches tend to be performed using search engines, while more open-ended searches tend to elicit hyperlink browsing. However, if there are no systematic relationships between the search goal and the users' cultural and linguistic backgrounds, the uncertainty about goals is likely to introduce unsystematic variation, which would strengthen rather than weaken the interpretability of the statistically validated impact of culture on search behaviour. We leave the question of whether the distribution of search goals differs systematically across countries and cultures to future research.

Possible links between culture and search goals might be the perception of the purpose of the Internet. The link between language and search goal might be found in language-related restrictions of the use of non-native language websites: since the cognitive effort is higher, the percentage of goal-oriented searches might be higher on L2 websites than on L1 websites.

Second, in study 3 we had to assume that the proportion of users with high vs. low domain knowledge who answered the questionnaire reflects the proportions in the website's total user population. Since this kind of information is not revealed by logfile data, we were not able to verify this assumption.

In study 3, we also had to make assumptions about the users' search purposes and resulting search strategy. In line with studies that found no differences between search strategies for familiar and unfamiliar topics (Kiestra, et al., 1994; Yee, 1993), we assumed that search strategies did not systematically differ between physicians and patients. In addition, the content-search of the website does not lend itself to exploratory search behaviour in the traditional sense. It can therefore be expected that most users' search purposes – independent of their domain knowledge – is more of a fact-finding than an exploratory nature (Allen, 1991), which strengthens our assumption.

Third, no insight is provided by logfile analysis about other personal and situational variables and their correlation with cultural and linguistic variables. With respect to culture, this particularly regards socio-economic variables (including Internet distribution and Internet experience). With respect to language, results from the supplementary questionnaire suggested a higher percentage of physicians among L2 users, which helped explain findings about search behaviour. Data analysis exclusively relying on logfile data would not have been able to reveal the role of domain knowledge.

Finally, due to its non-reactive nature, logfile analysis only allows for the collection of behavioural data. Attitudinal data cannot be assessed.

Nevertheless, investigation into navigation and search behaviour by means of logfiles taps authentic search histories, producing high ecological validity. As its common in other studies, it can be assumed that the size of the sample outweighs the noise.

In this context, it should also be mentioned that the quality of our results obtained in study 8 would increase if the crawler used, in particular the language identification feature – but also in terms of precision and recall[58], was validated by alternative crawlers (e.g. WIRE).

Questionnaires and Laboratory Studies. In studies where data was collected by means of an online survey or a laboratory study, we were able to assess this kind of information. However, online survey and laboratory studies generate other problems. This regards, first of all, a sampling bias through self-selection processes, in particular with regard to online surveys. In study 3, physicians might be underrepresented and patients over represented due to the fact that patients are more likely to visit that website during leisure time and, consequently, have more time to answer the questionnaire. Potential biasing self-selection effects in the same survey may also imply that only non-native speakers above a certain language proficiency threshold were assessed.

Furthermore, these reactive types of data collection usually attract participants that have a more extraverted personality, who exhibit, on average, decreased risk perception (e.g., Davison, et al., 2002). This bias regarded in particular study 5 (risk perception) and study 7 (attitudes towards data disclosure).

In the methodological part in the beginning of the thesis we mentioned the particular challenges within cross-cultural and cross-linguistic research regarding the measurement of the effects of language and culture. Aspects such as the inequality of constructs, methods, items, or in sampling were pointed. Sample bias through questionnaires is a particularly complex subject in cross-cultural research, since questions, answers, and scales need to be carefully translated and culturally adapted. These challenges were also encountered within our research and it appears that the questionnaires used were not sufficiently appropriate to overcome these challenges. The restricted consistency of outcomes, in particular in study 5, 6, and 9 raises the question of the suitability of the employed measures. Nevertheless, our results were partly in line with other studies (e.g. the minor/lack of an impact of culture on risk perception in study 5). Citing Karahanna, Evaristo, and Srite (2002), only minor explanatory value of culture from most cross-cultural studies result.

We therefore recommend paying particular attention to these challenges for future studies. It also remains up to future research to find out whether the low explanatory value of culture is due to

[58] "Precision measures how well the retrieved documents match the query, while recall indicates what fraction of the relevant documents are retrieved by the query." Pinkerton, B. (1994): Finding what People Want: Experiences with the WebCrawler, First International World-Wide Web Conference, Geneva, Switzerland.

methodological difficulties or whether the impact of culture is indeed minor. Results from our logfile analyses however suggested the first case.

The main problem of our laboratory studies consisted of the restricted resources. A completely randomized assignment in study 4 would have required (a) a higher number of participants per language and (b) an adapted version of the website for each language. This would have allowed us to test participants of each language on both an L1 and an L2 version. In a similar manner, cultural adaptations of the tested website (study 6) were not available, which was why no treatment assignment was possible.

Restriction to One Website. Another problem is the potential effect of the restriction of one site and domain that concerned all our analyses, independently of whether data was gathered through logfile analysis, online surveys, or laboratory studies. This is a common problem in website-usage studies. Resulting biases are numerous. For instance, it is possible that certain search option features and design features that are specific to our website (e.g., the pictorial display of content-organized search; the website is based on Western design) influence user behaviour. Also, the restriction to one specific website implied a restriction to one topic, namely e-Health. Studies 5, 6, and 9 in particular were therefore not able to draw conclusions about its applicability to other topic areas and other websites. The topic specificity also limited the selection of broadened aspects of culture to this area. In study 5, data about information gathering from a single information source was collected, due to the one website approach. Future research is necessary to provide more insight about the role of alternative information sources.

However, the extension to more than one site and domain is hardly attainable for large-scale field studies since it is very difficult to find a set of real-world, heavily frequented websites that differ only with respect to only one feature. We relied on the same website in our laboratory studies (studies 4, 9) and for our online surveys (studies 3, 5, 6, 7) for reasons of consistency and comparability.

Even though we were able to present results obtained from two different websites, data is still restricted to the medical domain and to a Western design. It is therefore not clear to which extent users' attitudes were affected by these content and design restrictions.

Limitations Specific to Cross-cultural and Cross-linguistic Research. The most frequent limitation of analyses that were based on data from online surveys or laboratory studies is the small sample size. Results from study 3, study 8, and study 9 should be interpreted with caution, due to

the low number of answer sets per language (study 8 and 9) or per analysed group (study 3). Their low number resulted in restricted possibilities for statistical calculations in study 8 and 9. A larger language sample permitting more powerful statistical analyses is needed in future research.

Our investigation in study 8 was furthermore limited to eight languages that are predominantly part of the indo-European language group. It is not clear how language similarity or cultural affiliation (due to the fact that these languages are mainly spoken within the Western culture) affected the results.

Analyses of the impact of language are also considerably challenged by the scarcity of Internet-related language statistics. Internet statistics regarding the number of webhosts are limited to a few languages and only regard the overall information per language. Furthermore, no data is obtainable that refers to the number of medical webpages. Despite a number of initiatives towards documenting the use of languages on the Internet (e.g. The Language Observatory Project - Mikami and Suzuki, 2004), reliable data isn't yet available. This shortcoming applies therefore to all language related studies in this field and definitely needs to be approached by future studies.

Also, in order to evaluate language-related aspects, a more potent language identifier[59] is necessary. Language identification is particularly important due to the fact that correspondence between languages and country-level-domains are not always given. The small overlap between results from the web crawler and those from the server-log points out the existing technical limitations and need for technical developments. Due to this problem, language is often inferred from the country, which of course only represents a vague proxy.

A number of methodological difficulties particular to quantitative research were pointed out in the first chapter of this thesis. In order to overcome some of these difficulties, in particular problems of causality, a multi-nation approach and a broadened concept of culture were applied. Nevertheless, problems of causality were also found in our studies. Concerning study 6, the potential arbitrary nature of the variables selected was mentioned. Furthermore, for the purpose of our studies we incorporated the variables that were expected to have an impact. In study 7, we assumed that attitude towards data disclosure was solely influenced by the tested cultural variables and not by (potentially) correlating variables or that not assessed variables were equally distributed

[59] In particular, for Spanish the language recognizer seems to produce many errors by identifying Catalan.

among the groups.However, we have no insight into which other culture-related variables determine the users' attitude with regard to the aspects investigated.

Similar problems of correlations with other variables also apply to the investigation of language. For example, in study 9 no data about the users' (Internet/computer) literacy levels or domain knowledge was collected. These variables may determine the value of a website and hence the probability of accessing the website. As shown in chapter 2, domain knowledge enhances language proficiency. It may therefore also affect the user's perceived costs. Also, the comparison of the perception of native language offer with the actual Internet offer was based on the assumption that native language information is perceived as equally interesting and important to each group of native speakers. For example, if Chinese native speakers perceive Chinese (Mandarin) online information as not sufficient despite a larger number of webhosts, this might be due to the low relevance of the offered information in Chinese language.

In study 8, Russian, despite being an L2 language, showed similarities with the L1 language. The consistency indicates a correlating variable, such as cultural affiliation. Since we did not assess further data, an explanation of the findings remains up to future research. Cultural affiliation might also be the reason that affected the number of German native speakers visiting the website and the high number of in-links from German-language websites. Furthermore, some languages are spoken only within one country or within a larger region with homogeneous technological development stages. Results from recent research indicate that connectivity between websites is also influenced by the age of the Internet in a region (referenz!). Websites from countries/regions where the Internet has been adopted only recently are therefore less likely to be linked to our investigated website. Our results might therefore be biased to a certain degree by a correlation between language and the year of Internet adoption. Although these effects were partly controlled taking the number of Internet users into account (which should increase with a higher number of years since Internet adoption) and having early and later adopters in both language groups, future research is needed to gain more insight.Finally, due to the fact that culture and language are personal variables, i.e. variables that are inherent to the participants, certain research designs are excluded or difficult to attain. This problem is increased by the restricted availability of language versions of the tested website, as well as by the invariance of the design. This matter was addressed in section 1.3.3. Accordingly, our empirical work is mainly based on ex post investigation/correlative studies. Due to the difficult manipulation of the independent variable, results from this kind of investigation are technically limited to findings about correlations and/or can only infer findings about causality with caution.

5.3 THESIS CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

This thesis investigated the impact of culture and language on the use of Information Systems, in particular websites, their services and products. The impact was tested in a number of empirical studies that were all based on a quantitative approach. The studies were applied to health care, which is one of the most important emerging topic areas in the broader field of Information Systems. In contrast to the vast majority of cross-cultural and cross-linguistic studies, we gathered data through logfile analysis and web-crawling, which are – despite their advantages – rarely found in cross-cultural and cross-linguistic studies.

The role of language is to a certain extent intertwined with that of culture. The investigation of the one subject can therefore hardly be separated from the other. Yet even though language is in most cross-cultural studies considered an overt characteristic of culture, we presented argumentations why language should also be examined as an independent variable.

Our empirical studies were preceded by a thorough conceptual discussion of the two investigated variables: culture and language (chapter 1). The subject of culture was mainly discussed with respect to its methodological and conceptual challenges, also resulting from its difficult position between social sciences and quantitative approaches that comply with the needs of more application-oriented sciences such as Information Systems. Despite existing criticism of Hofstede's paradigm, we based our quantitative research mainly on his data. One reason is the lack of alternative quantitative data, but we also presented arguments in favour of his approach within the context of Information Systems. We furthermore suggested a broadened approach to culture that integrates more dynamic aspects of society such as socio-economic variables or individuals' knowledge. These broadened approaches were applied in studies where users' attitudes were expected to be influenced by these aspects (chapter 3).

With respect to language, two main perspectives were presented: the user perspective and the provider perspective. Regarding the user perspective, we examined the impact of language on individuals' cognitive effort and aspects of attitudes towards language(s). With respect to the provider perspective, language was investigated from a market point of view, mainly regarding language as a marketing tool for product differentiation. For the purpose of the thesis, the role of language was reduced in our empirical studies to an opposition of native speakers and non-native speakers.

Findings from our studies revealed the rather strong impact of language and culture on behavioural aspects of the users of Information Systems. They furthermore provided evidence for the impact on attitudinal aspects that however need to be looked at more in detail in future research.

The Impact of Culture. It was found that users' cultural backgrounds significantly affected navigational behaviour in general, and search behaviour in particular. Evidence was provided for the role of culture as a determinant of the user's information need (Uncertainty Avoidance, context specificity), his/her attitude towards time (Long-term Orientation), perception of space and hierarchical levels (Power Distance), and the structuring of time and behavioural steps (Monochronicity). These culturally determined differences result in differences regarding the amount of page sets visited, the time that is spent on page visits, as well as the preference for search engines, the use of hyperlinks, and the linearity of navigational patterns. These findings extend previous research as they provide evidence for culturally determined website navigation *behaviour* as a counterpart of culturally determined website *design* (Marcus and West Gould, 2000). As a consequence of the behavioural regularities discovered, we also suggested a separation into two main cultural groups with two complementary design suggestion sets.

The role of culture with regard to information need, expressed through risk perception and product valuation, seems to be less salient. This outcome confirms previous research that found a minor impact of culture on risk and trust perception (Jarvenpaa and Tractinsky, 1999; Ko, et al., 2004). However, results also suggest that the context of online health *information services* represent a low involvement situation that triggers less attitudinal differences between users of different cultural backgrounds. It appears that the cultural dimension of Uncertainty Avoidance is the most important determinant of a user's perception of information need. This result is in line with the role of Uncertainty Avoidance as a determinant of the amount of information needed, obtained in chapter 2. However, results from both chapters indicate the rather complex impact of this specific cultural dimension. It seems that users from low uncertainty avoidant cultures have a higher need for additional information, whereas users of high uncertainty avoidant cultures have a higher need for accessing existing information. In isolated cases we also found evidence of the impact of other cultural dimensions. Future research is clearly needed here.

Furthermore, results suggest that the chosen aspects of a broadened concept of culture (perceived quality of health care and domain knowledge) have a more consistent impact than stable cultural values. Yet, we also mentioned the difficult separation of broadening cultural variables and control variables where the distinction predominantly underlies the researcher's perspective.

Results regarding users' attitudes towards data disclosure clearly show the impact of culture. Hence they question the generalizability of previous results in this field. In contrast to previous studies, we argued that both Individualism *and* Power Distance affect privacy attitudes and can therefore not be considered independently from each other. Empirical results provided evidence that the combination of both cultural dimensions has in fact the strongest impact.

The Impact of Language. With regard to the impact of language on users' cognitive effort, we investigated the extent to which the use of search options is affected by the user's language skills. As a result of our studies' outcomes, our expectations of differences between native speakers and non-native speakers and their use of alphabetical searches and search engines needed to be restricted. It was found that only non-native speakers with low domain knowledge differ from native speakers in their use of search options. No significant differences between non-native speakers with high domain knowledge and native speakers were revealed. Confirming these results, study 4 found that the use of the user's native language leads low domain knowledge users to a higher usefulness and ease of use. These outcomes suggest that the user's domain knowledge – in many cases linked to the user's education, plays an - at least - equally central role.

Further evidence for the important role of language and discrepancies between native speakers and non-native speakers were found with respect to website access. Websites are visited by a lower percentage of L2 users than one would expect from the respective proportion of Internet users. Reasons for this are the higher cognitive effort (as a consequence, L2 users are represented by a higher percentage of high domain knowledge users than L1 users) and the a lower percentage of in-links. This is the outcome of the investigation of the impact of language on link setting behaviour and link following behaviour described in study 8. In this combined form, the investigation is an important contribution towards insight about the role of language. Outcomes emphasize the double disadvantage for lesser-used languages with regard to information access: first, links to non-native languages are less likely to be followed, and second, lesser-used languages have a lower number of webhosts and a lower number of hyperlinks. The impact of language on hyperlink distribution also shows the significance of language as a general border for information flow on the Internet, which becomes increasingly important when legal and logistical barriers disappear.

Similar to our cross-cultural investigations, the impact of language on attitudinal variables is less clear and needs future research. Interestingly, results from study 9 suggest that L1 websites generate higher satisfaction for user groups that are not accustomed to be provided with native language information. A focus on these user groups in a website's language offering would comply with

Grin's "discriminatory language market goals" (Grin, 1994). For L2 users, website satisfaction depends mainly on language proficiency levels. Further studies are clearly encouraged given the importance of the topic.

Suggestions for Future Research. Suggestions for Future Research result predominantly from the limitations of our studies as presented above.

The most important challenges encountered in our research are of a methodological nature, partly resulting from resource restrictions. These are especially present in cross-cultural research, and can only be overcome given the current state of the art with a major time and financial investment, if at all.

Future research should first of all concentrate on investigating how website preferences – determined by culture, language, and education - can be met with tailored website design. In order to allow for a valid experiment design, studies with adapted websites are necessary, both before and after the adaptation. With regard to language this means that native speakers of the same language should be tested with an L1 and L2 version. With regard to culture, members of each cultural group should be tested with a culturally adapated website version and a non-adapted version. Table 9 presented several features that are potentially subject to adaptation. The selected features were chosen as a result of our findings from behavioural analyses. It remains up to further studies to investigate the degree to which these behavioural preferences are linked to attitudinal aspects. Knowledge about this issue is helpful in determining the role of culturally adapted website design for generating higher satisfaction. Furthermore, future research should establish whether behavioural preferences are a conscious decision and correspond to the perceived effectiveness and efficiency of the respective feature.

As for the role of language and domain knowledge, synonym expansion and spell checking should be considered where appropriate. Since we only focussed on selected independent and dependent variables, future studies should take additional cultural dimensions, other aspects of website design, and additional search options and search cues into account, as well as analyse search words for detecting culturally and linguistically determined patterns of information seeking. Further topic areas also have to be investigated, and user behaviour needs to be examined with respect to more than one website.

Furthermore, cognitive processes and *"perceived ease of use"* on the one hand and *"perceived usefulness"* and attitudinal processes on the other hand were mainly investigated independently from each

other. Research resulting from previous studies suggest varying interaction between the two aspects (see section 5.1.1). Future cross-cultural and cross-linguistic investigation should therefore carry out further studies that determine the interaction of cognitive and attitudinal processes, of *"perceived ease of use"* and *"perceived usefulness"* in more detail, within a cultural and linguistic context.

The combination of logfile analysis and online surveys appears to be an appropriate methodological combination for cross-cultural and cross-linguistic investigations, whose application for future research is hereby encouraged. Insight from future combined analyses would provide, for example, useful knowledge about cultural or language biases from self-selection in questionnaires. In our research, this combination was used to verify the geographical information inferred from IP addresses. Findings from the pilot test in study 4 suggest employing think-aloud methods for gaining more detailed insight. In most studies, the analysis of language was reduced to a comparison of L1 and L2 users. Future research that investigates the impact of proficiency levels in non-native languages as well as literacy levels in native languages is encouraged. As a result of our studies, the question of where the language proficiency threshold below which non-native language websites are not accessed is raised.

We also want to encourage carrying out further research that finds explanations for our findings from exploratory parts of our empirical investigations.

Finally, the studies in this thesis were based on the assumption that the effects of culture and language are independent from each other. Nevertheless, there is little empirical evidence (e.g., Luna, et al., 2003) in favour nor against this assumption. It would therefore be extremely beneficial to gain knowledge about the interaction of culture and language, potential mediation effects, and their combined impact on Internet users' behaviour and attitudes.

6. REFERENCES

Languages and Internet, UNESCO Culture Sector, 2005, October 10, UNESCO Culture Sector, http://portal.unesco.org/culture/en/ev.php-URL_ID=21296&URL_DO=DO_TOPIC&URL_SECTION=201.html

(2005): EHealth. Developing a Methodological Framework for High-Quality Assessment of the IST-RTD Effects (Results and Impacts) at the "Strategic Objective" Level, European Commission, March, MIST Plus Interim Report, Commission Contract No 29000, European Commission, http://europa.eu.int/comm/dgs/information_society/evaluation/data/pdf/studies/2004_indicators_d02_ehealth.pdf

(2005): Eurobarometer: Europeans and Languages, European Commission, Special Eurobarometer 237 - Waver 63.4 - TNS Opinion & Social, European Commission, http://europa.eu.int/comm/public_opinion/archives/ebs/ebs_237.en.pdf

Adamic, L.A. (1999): The Small World Web, Third European Conference of Research and Advanced Technology for Digital Libraries (ECDL '99), Paris, France. URL: http://www.hpl.hp.com/shl/papers/smallworld/smallworldpaper.html.

Adamic, L.A. and Adar, E. (2001): You Are What You Link., Tenth International World Wide Web Conference (WWW 2001), Hong Kong. URL: http://www10.org/program/society/yawyl/YouAreWhatYouLink.htm.

Adler, N. (1991): International Dimension of Organizational Behavior, POWs-Kent, Boston.

Agourram, H. and Saucier, M. (2004): The Use of Open and Repertory Grid Interview Techniques in Cross-Cultural Research, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.

Ajzen, I. and Fishbein, M. (1977): Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research, Psychological Bulletin 84, pp. 888-918.

Ajzen, I. and Fishbein, M. (1980): Understanding Attitudes and Predicting Social Behavior, Prentice-Hall, Englewood Cliffs, NJ.

Akerlof, G. (1984): An Economic Theorist's Book of Tales, Cambridge University Press, Cambridge.

Albert, R.; Jeong, H. and Barabasi, A.-L. (1999): Diameter of the World Wide Web, Nature 401 [9], pp. 130-131.

Albrecht, G. (1982): Zur Stellung historischer Forschungsmethoden und nichtreaktiver Methoden im System der empirischen Sozialforschung, Ludz, P.C., Soziologie und Sozialgeschichte (Sonderheft 16 der Kölner Zeitschrift für Soziologie und Sozialpsychologie) pp. 242-293, Westdeutscher Verlag, Opladen.

Albrecht, G. (1985): Nichtreaktive Messung und Anwendung historischer Methoden, van Wieken-Meiser, K.J., Techniken der empirischen Sozialforschung pp. 9-81, Oldenbourgh, München, Wien.

Allen, B.L. (1991): Cognitive Research in Information Science: Implication for Design, Annual Review of Information Science and Technology 26, pp. 3-37.

Allinson, L. and Hammond, R. (1989): A Learning Support Environment: The Hitchhiker's Guide, McAleese, R., Hypertext: Theory into Practice pp. 62-74, Ablex, Norwood, NJ.

Allport, G.W. (1935): Attitudes, Murchison, C., Handbook of Social Psychology pp. 798-844, Clark University Press, Worcester, MA.

Allport, G.W.; Vernon, P.E. and Lindzey, G. (1931): Study of Values: A Scale for Measuring the Dominant Interests in Personality, Houghton Mifflin, Boston.

Anderson, J.R and Bower, G.H. (1973): Human Associative Memory, Wiley & Sons, New York.

Arcand, J.-L. (1996): Development in Economics and Language: The Earnest Search for a Mirage, International Journal of the Sociology of Language 121, pp. 119-157.
ATIS (2001): National Information Systems Security (INFOSEC) Glossary, 2005, August 3, http://www.atis.org/tg2k/_information_system.html
Baeza-Yates, R. and Castillo, C. (2001): Relating Web Characteristics with Link Based Web Page Raking., Eighth International Symposium on String Processing and Information Retrieval (SPIRE 2001), IEEE, Laguna San Rafael, Chile.
Baeza-Yates, R. and Poblete, B. (2003): Evolution of the Web Structure, 12th International World Wide Web Conference (WWW 2003) (Poster), Budapest, Hungary. URL: http://www2003.org/cdrom/papers/poster/p103/p103-baeza-yates/p103-baeza-yates.html
Barnett, G. (2003): The Structure of International Internet Flows, Internet Researcher 4.0 (AoIR), Toronto, Canada.
Baumgartner, V.-J. (2003): A Practical Set of Cultural Dimensions for Global User-Interface Analysis and Design, Diplomarbeit (Master Thesis), Informations-Design, FH Joanneum, Graz. URL: http://mavas.at/val/downloads/ValBaumgartner_PracticalSetOfCulturalDimensions.pdf
Bechwati, N.N. and Xia, L. (2003): Do Computers Sweat? The Impact of Perceived Effort of Online Decision Aids on Consumers' Satisfaction with the Decision Process, Journal of Consumer Psychology 13 [1&2], pp. 139-148.
Berendt, B. and Brenstein, E. (2001): Visualizing Individual Differences in Web Navigation: STRATDYN, a Tool for Analyzing Navigation Patterns, Behavior Research Methods, Instruments & Computers 33, pp. 243-257.
Berendt, B.; Mobasher, B.; Nakagawa, M. and Spiliopoulou, M. (2002): The Impact of Site Structure and User Environment on Session Reconcstruction in Web Usage Analysis, WEBKDD 2002 - Mining Web Data for Discovering Usage Patterns and Profiles, Edmonton, Canada.
Berendt, B.; Mobasher, B.; Spiliopoulou, M. and Wiltshire, J. (2001): Measuring the Accuracy of Sessionizers for Web Usage Analysis, Workshop on Web Mining at the First SIAM International Conference on Data Mining, Chicago, IL, USA. URL: http://www.hhl.de/fileadmin/LS/micro/Download/Berendt_2001_MeasuringTheAccuracy.pdf
Bernard, M.L. (2000): Examining a Metric for Predicting the Accessibility of Information within a Hypertext, Ph.D. Thesis, Human Factors Psychology, Wichita State University. URL: http://psychology.wichita.edu/mbernard/Dissertation.pub..pdf
Berry, J.W.; Portinga, H.Y.; Segall, M.H. and Dasen, P.R. (1992): Cross-Cultural Psychology: Research and Applications, Cambridge University Press, Cambridge, UK.
Berry, J.W.; Portinga, H.Y.; Segall, M.H. and Dasen, P.R. (2002): Cross-Cultural Psychology, University Press, Cambridge.
Berry, M.J. and Linoff, G. (1997): Data Mining Techniques: For Marketing, Sales, and Customer Support, John Wiley & Sons, New York, NY.
Bharat, K.; Chang, B.-W.; Henziger, M. and Ruhl, M. (2001): Who Links to Whom: Mining Linkage betwen Web Sites, IEEE International Conference on Data Mining (ICDM '01), San Jose, CA, USA.
Blake, E. (2001): A Field Computer for Animal Trackers, 2nd South African Conference on Human-Computer Interaction (CHI-SA 2001), Pretoria, South Africa. URL: http://www.chi-sa.org.za/CHI-SA2001/proceedings/cybertracker.pdf.
Blank, A. (2001): Einführung in die lexikalische Semantik, Gunter Narr, Tübingen.
Bodley, J. (2000): Cultural Anthropology: Tribes, States, and the Global System, 3. ed., Mayfield Publishing Co., Mountain View, CA.
Bortz, J. (2004): Statistik für Human- und Sozialwissenschaftler, Springer, Berlin.

Botafogo, R.A.; Rivlin, E. and Shneiderman, B. (1992): Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics, ACM Transactions on Information Systems 10 [2], pp. 142-180.
Bourges-Waldegg, P. and Scrivener, S.A.R. (1998): Meaning, the Central Issue in Cross-Cultural HCI Design, Interacting with Computers 9 [3], pp. 287-309.
Brachinger, H.W. and Weber, M. (1997): Risk as a Primitive - A Survey of Measures of Perceived Risk, Operations Research - Spektrum 19 [3], pp. 235-250.
Bracht, G.H. and Glass, G.V. (1968): The External Validity of Experiments, American Education Research Journal 5, pp. 437-474.
Branting, L.K. (2004): Learning Features Weights from Customer Return-Set Selections, Knowledge and Information Systems 6 [2], pp. 188-202.
Breton, A. (1998): Economic Approaches to Language and Bilingualism, Canadian Heritage, Ottawa.
Briley, D.A.; Morris, M.W. and Simonson, I. (2002): Reasons as Carriers of Culture: Dynamic versus Dispositional Models of Cultural Influence on Decision Making, Journal of Consumer Research 27, pp. 157-177.
Brin, S. and Page, L. (1998): The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Networks 30 [1-7], pp. 107-117.
Brislin, R. (1986): The Wording and Translation of Research Instruments, Lonner, W.J. and Berry, J.W., Field Methods in Cross-Cultural Research pp. 1337-1364, Sage, Newbury Park, CA.
Broder, A.; Kumar, R.; Maghoul, F.; Raghaven, P.; Stata, R.; Tomkins, A. and Wiener, J. (2000): Graph Structure of the Web, Ninth International World Wide Web Conference (WWW 2000), Amsterdam, The Netherlands. URL: http://www.almaden.ibm.com/cs/k53/www9.final/
Bronold, R. (1999): Mediengerechte Online-Forschung: Das GfK Online-Forschungsprogramm, Batinic, B.; Gräf, L.; Werner, A. and Bandilla, W., Online Research. Methoden, Anwendungen und Ergebnisse pp. 39-45, Hogrefe, Göttingen.
Brunn, S. and Dodge, M. (2001): Mapping the "Worlds" of the World-Wide Web: Restructuring Global Commerce through Hyperlink., American Behavioral Scientist 44 [10], pp. 1717-1739.
Bryan, N.B.; McLean, E.R.; Smith, S.J. and Burn, J. (1994): The Structure of Work Perceptions among Hong Kong and United States IS Professionals: a Multidimensional Scaling Test of the Hofstede Cultural Paradigm, Computer Personnel Research Conference on Reinventing IS: Managing Information Technology in Changing Organizations, Alexandria, Virginia, USA.
Buckley, G. (2001): Semantics, http://www.ling.upenn.edu/courses/Spring_2001/ling001/semantics.html
Bungard, W. and Lück, H.E. (1974): Forschungsartefakte und nicht-reaktive Meßverfahren, Teubner, Stuttgart.
Cadez, I.; Heckerman, D. and Meek, C. (2000): Visualization of Navigation Patterns on a Web Site Using Model Based Clustering, Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, MA, USA.
Campbell, D.T.; Cook, T.D. and Shadish, W.R. (2002): Experimental and Quasi-Experimental Designs for Generalized Causal Interference, Houghton-Mifflin, New York.
Canter, R.; Rivers, R. and Storrs, G. (1985): Characterizing User Navigation through Complex Data Structures, Behaviour and Information Technology 4 [2], pp. 93-102.
Carey, T.T.; Peerenboom, D.S.; Mitchell, S. and Lytwyn, M. (1998): Design Evolution in a Multimedia Tutorial on User-Centered Design, ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'98), Los Angeles, CA, USA.
Carrell, P.L. and Wise, T.E. (1998): The Relationship between Prior Topic Knowledge and Topic Interest in Second Language Reading, Studies in Second Language Acquisition 20, pp. 285-309.
Case, D.O. (2002): Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior, Academic Press, New York.

Cavnar, W.B. and Trenkle, J.M. (1994): N-Gram-Based Text Categorization, Third Annual Symposium on Document Analysis and Information Retrieval (SDAIR-94), Las Vegas, NE, USA.
Chen, C. and Rada, R. (1996): Interacting with Hypertext: a Meta-Analysis of Experimental Studies, Human-Computer Interaction 11, pp. 125-156.
Chimera, R. and Shneiderman, B. (1994): An Exploratory Evaluation of Three Interfaces for Browsing Large Hierarchical Tables of Contents, ACM Transactions on Information Systems 12 [4], pp. 383-406.
Chiswick, B.R. and Miller, P.W. (1995): The Endogeneity between Language and Earnings: International Analyses, Journal of Labor Economics 13, pp. 246-288.
Chomsky, N. (1957): Syntactic Structures, Mouton, The Hague, Netherlands.
Ciborowski, I.J. (1979): Cross-Cultural Aspects of Cognitive Functioning: Culture and Knowledge, Marsella, A.J.; Tharp, R.G. and Ciberowski, T.J., Perspectives in Cross-Cultural Psychology pp. 101-116, Academic Press, New York.
Cilliers, P (1998): Complexity and Postmodernism: Understanding Complex Systems, Routledge, London.
Clarke, R. (2000): Appropriate Research Methods for Electronic Commerce, 2005, October 1, 2000, April 19, http://www.anu.edu.au/people/Roger.Clarke/EC/ResMeth.html
Coady, M.R. (2001): Attitudes towards Bilingualism in Ireland, Bilingual Research Journal 25 [1&2], pp. 39-58.
Cohen, J. (1968): Multiple Regression as General Data-analytic System, Psychological Bulletin 70, pp. 426-443.
Collins, A.M. and Quilian, M.R. (1969): Retrieval Time from Semantic Memory, Journal of Verbal Learning and Verbal Behavior 8, pp. 240-248.
Conklin, E.J. (1987): Hypertext: An Introduction and Survey, IEEE Computer 20 [9], pp. 17-41.
Cooley, R.; Mobasher, B. and Srivastava, J. (1999): Data Preparation for Mining World Wide Web Browsing Patterns, Journal of Knowledge and Information Systems 1 [1], pp. 5-32.
Cox, D.F. (1967): Risk Handling in Consumer Behaviour, Cox, D.F., Risk Handling in Consumer Behaviour - an Intensive Study of Two Cases pp. 34-81, Harvard University, Boston, MA.
Crystal, D. (1992): An Encyclopedic Dictionary of Language and Languages, Blackwell, Cambridge, MA.
Cunliffe, D.; Taylor, C. and Tudhope, D. (1997): Query-based Navigation in Semantically Indexed Hypermedia, Eighth ACM Conference on Hypertext, Southhampton, UK.
Cunningham, S. (1967): The Major Dimensions of Perceived Risk, Cox, D.F., Risk Taking and Information Handling in Consumer Behaviour pp. 82-108, Cambridge, MA: Harvard University Press.
Cyr, D.; Bonanni, C. and lsever, J. (2004): Design and E-Loyalty across Cultures in Electronic Commerce, Sixth International Conference on Electronic Commerce (ICEC04), Delft, The Netherlands.
Cyr, D.; Ilsever, J.; Bonanni, C. and Bowes, J. (2004): Website Design and Culture: An Empirical Investigation, Sixth International Workshop on Internationalisation of Products and Systems (IWIPS), Vancouver, Canada.
Dafoulas, G. and Macaulay, L. (2001): Investigating Cultural Differences in Software Teams, Electronice Journal on Information Systems in Developing Countries 7 [4], pp. 1-14.
Dahl, S. (2004): Intercultural Research: The Current State of Knowledge, 2005, October 1, Middlesex University Discussion Paper No. 26, January 12, 2004, Discussion Paper No. 26, http://ssrn.com/abstract=658202
Dalmazzone, S. (1999): Economics of Language: A Network Externalities Approach, Breton, A., Exploring the Economics of Language. New Canadian Perspectives pp. 63-87, Canadian Heritage, Ottawa.

Danielson, D.R. (2002): Transitional Volatility in Web Navigation: Usability Metrics and User Behavior., M.S. Thesis, Symbolic Systems Program, Stanford University. URL: http://www.stanford.edu/~davidd/MastersThesis/Danielson-Transitional_Volatility_in_Web_Navigation.pdf
Danowski, J. and Edison-Swift, P. (1985): Crisis Effects on Intraorganizational Computer-Based Communication, Communication Research 12 [2], pp. 251-270.
Darlington, G. (1996): Culture - a Theoretical Review, Joynt, P. and Warner, M., Managing across Cultures: Issues and Perspectives, International Thompson, London.
Davies, S. (1996): Big Brother: Britain Web of Surveillance and the New Technological Order, Pan, London.
Davis, F.D. (1989): Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, MIS Quarterly 13, pp. 319-340.
Davis, F.D. (1993): User Acceptance of Information Technology: System Characteristics, User Perceptions and Behavioural Impacts, International Journal of Man-Machine Studies 38, pp. 475-487.
Davison, R.M.; Clarke, R.; Kuo, B.; Langford, D. and Smith, H.J. (2002): Information Privacy in a Globally Networked Society: Implications for IS Research, 23rd International Conference on Information Systems (Panel Session) (ICIS), Barcelona, Spain. URL: http://www.is.cityu.edu.hk/research/resources/icis2002.htm [last visited: September 28, 2005]
Davison, R.M.; Clarke, R.; Smith, H.J.; Langford, D. and Kuo, B. (2003): Information Privacy in a Globally Networked Society: Implications for IS Research, Communications of the Association for Information Systems 12, pp. 341-365.
Dawes, R.M. (1972): Fundamentals of Attitude Measurement, John Wiley and Sons, New York, NY.
De la Fuente, M.J. (2002): Negotiation and Oral Acquisition of L2 Vocabulary, Studies in Second Language Acquisition 22, pp. 81-112.
De Mooij, Marieke (2003): Convergence and Divergence in Consumer Behaviour: Implications for Global Advertising, International Journal of Advertising 22 [2], pp. 183-202.
De Souza, M. and Dejean, P. (1999): Interculturality and Design: Is Culture a Block or an Encouragment to Innovation., International Conference of Design Research: Design Cultures, Sheffield, UK. URL: http://www.shu.ac.uk/schools/cs/ead/work/desouz.pdf
Deaux, K. and Martin, D. (2003): Interpersonal Networks and Social Categories. Specifying Levels of Context in Identity Processes, Social Psychology Quarterly 66, pp. 101-117.
Del Galdo, E.M. and Nielsen, J. (1996): International User Interfaces, John Wiley & Sons, New York, NY.
DePaula, R. (2003): A New Era in Human Computer Interaction: The Challenges of Technology as Social Proxy, Latin American Conference on Human-Computer Interaction, Rio de Janeiro, Brazil.
Dholakia, U. and Bagozzi, R. (2001): Consumer Behaviour in Digital Environments, Wind, J. and Mahajan, V., Digital Marketing pp. 163-200, Wiley, New York.
Dickson, P. (2000): Understanding the Trade Winds: The Global Evolution of Production, Consumption, and the Internet, Journal of Consumer Research 27 [1], pp. 115-122.
Dirven, R. and Verspoor, M. (1998): Cognitive Exploration of Language and Linguistics, John Benjamins, Amsterdam, Philadelphia.
Dmoch, T. (1997): Interkulturelle Werbung. Verhaltenswissenschaftliche Grundlage für die Standardisierung erlebnisbetonter Werbung, Shaker, Aachen.
Dorian, N.C. (1981): Language Death: The Life Cycle of a Scottish Gaelic Dialect, University of Pennsylvania Press, Philadelphia.
Dowling, G. and Staelin, R. (1994): A Model of Perceived Risk and Intended Risk-Handling Activity, Journal of Consumer Research 21, pp. 119-134.

Drinkwater, S.J. and O'Leary, N.C. (1997): Unemployment in Wales: Does Language Matter?, Regional Studies 31, pp. 583-591.
Dufour, R. and Kroll, J. (1995): Matching Words to Concepts in Two Languages. A Test of the Concept Mediation Model of Bilingual Representation, Memory & Cognition 23, pp. 166-180.
Dufresne, A. and Turcotte, S. (1997): Cognitive Style and its Implications for Navigation Strategies, Boulay, B. and Mizgouchi, R., Artificial Intelligence in Education: Knowledge and Media in Learning Systems (Proceedings of AI ED '97 World Conference on Artificial Intelligence Education) pp. 287-293, IOS Press, Amsterdam.
Duncker, E. (2002): Cross-Cultural Usability of the Library Metaphor, ACM & IEEE Joint Conference on Digital Libraries (JCDL '02), Portland, OR, USA.
Dunckley, L. and Kuldip, J. (2004): Formative Evaluation of International User Interfaces Based on Sorting Techniques, Sixth International Workshop on Internationalisation of Products and Systems (IWIPS), Vancouver, Canada.
Duranti, Alessandro (1997): Linguistic Anthropology, Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge.
Ellis, D. (1989): A Behavioural Approach to Information Retrieval System Design, Journal of Documentation 45 [3], pp. 171-212.
Evans, M.; McBride, A.; Queen, M.; Thayer, A. and Spyridakis, J. (2004): Has the Tone of English Become Globalized? An Empirical Research Study Investigating the Written Tone of University Websites around the World, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.
Evaristo, J.R. (2004): Cross-Cultural Research in MIS, ISWorld Encyclopedia, 2005, October 1, 2004, August 27, http://ispedia.terry.uga.edu/?title=Cross-Cultural_Research_in_MIS
Evaristo, J.R. and McCubbrey, D. (1995): Editorial: Special Issue on Cross Cultural Management in Information Systems, Information Technology and People 8 [2], pp. 3-5.
Evers, V. and Day, D. (1997): The Role of Culture in Interface Acceptance, Human Computer Interaction (INTERACT '97), London, UK.
Evers, V.; Kukulska-Hulme, A.H. and Jones, A.C. (2001): Cross-Cultural Understanding of Graphical Elements on the DirectED website, Annual Workshop on Cultural Issues on HCI, Putteridge Bury, UK. URL: http://staff.science.uva.nl/~evers/pubs/cult%20issues%20luton%20ed.pdf
Fishman, J.A. (1999): Handbok of Language and Ethnic Identity, Oxford University Press, Oxford.
Fishman, J.A. (2001): Can Threatened Languages be Saved?, Multilingual Matters Ltd., Clevedon.
Fitzgerald, W. (2004): Models for Cross-Cultural Communications for Cross-Cultural Website Design, National Research Council Canada, 11, 2004, April 6, NRC/ERB-1108, NRC 46563, http://iit-iti.nrc-cnrc.gc.ca/publications/nrc-46563_e.html
Ford, G.; Kotzè, P. and Marcus, A. (2005): Cultural Dimensions: Who is Stereotyping Whom?, 11th International Conference on Human-Computer Interaction (HCI), Las Vegas, NE, USA.
Freeman, L.C. (1979): Centrality in Social Networks: Conceptual Clarification, Social Networks 1, pp. 215-239.
Gaenslen, F. (1986): Culture and Decision Making in China, Japan, Russia, and the United States, World Politics 39, pp. 87-103.
Gardner, R.C. (1985): Social Psychology and Second Language Learning: The Role of Attitudes and Motivation, Edward Arnold, Bungay, Suffolk.
Geerarts, D. (2003): Cultural Models of Linguistic Standardization, Dirven, R.; Roslyn, F. and Pütz, M., Cognitive Models in Language and Thought. Ideology, Metaphors and Meanings pp. 25-68, Mouton de Gruyter, Berlin.
Gemünden, H.G. (1985): Perceived Risk and Information Search. A Systematic Meta-Analysis of the Empirical Evidence, International Journal of Research in Marketing 2 [2], pp. 79-100.
Giles, H. and Johnson, P. (1981): The Role of Language in Ethnic Group Formation, Turner, J.C. and Giles, H., Intergroup Behaviour pp. 199-243, Blackwell, Oxford.

Giles, H. and Byrne, J.L. (1982): An Intergroup Approach to Second Language Acquisition, Multilingual and Multicultural Development 3, pp. 17-40.
Gillham, R. (2005): Diary Studies as a Tool for Efficient Cross-Cultural Design, Seventh International Workshop on Internationalisation of Products and Systems (IWIPS), Amsterdam, The Netherlands.
Goldman, S.R. and Duran, R.P. (1988): Answering Questions from Oceanography Texts: Learner, Task, and Text Characteristics, Discourse Processes 1, pp. 373-412.
Goodwin, N.C. (1987): Functionality and Usability, Comunications of the ACM 30, pp. 229-233.
Gould, J.D.; Boies, S.J: and Lewis, C. (1991): Making Usable, Useful, Productivity-enhancing Computer Applications, Communications of the ACM 34, pp. 74-85.
Grin, F. (1994): The Economics of Language. Match of Mismatch?, Revue Internationale de Science Politique 15 [1], pp. 25-42.
Grin, F. (1996): The Economics of Language: Survey, Assessment, and Prospects, International Journal of Society and Language 121, pp. 17-44.
Gudykunst, W.B.; Matsumoto, Y.; Ting-Toomey; S. Nishida, T.; Kim, K. and Heyman, S. (1996): The Influence of Cultural Individualism-Collectivism, Self-Construals, and Individual Values on Communication Styles across Cultures, Human Communication Research 22 [4], pp. 510-543.
Gudykunst, W.B. and Ting-Toomey, S. (1988): Culture and Interpersonal Communication, Sage, Newbury Park, CA.
Gudykunst, W.B.; Ting-Toomey, S.; Hall, B.J. and Schmidt, K.L. (1989): Language and Intergroup Communication, Asante, M.K. and Gudykunst, W.B., Handbook of International and Intercultural Communication pp. 145-158, Sage, Newbury Park, London, New Delhi.
Gujarati, D.N. (2003): Basic Econometrics, 4th. ed., McGraw-Hill, Boston, MA.
Gürhan-Canli, Z. and Maheswaran, D. (2000): Determinants of Country-of-Origin Evaluations, Journal of Consumer Research: An Interdisciplinary Quarterly 27 [1], pp. 96-108.
Hahne, A. (2001): What's Different in Second Language Processing? Evidence from Event-Related Brain Potentials, Journal of Psycholinguistic Research 30, pp. 199-212.
Halavais, A. (2000): National Borders on the World Wide Web, New Media and Society 2 [1], pp. 7-28.
Hall, B. and Gudykunst, W. (1986): The Intergroup Theory of Second Language Ability, Journal of Language and Social Psychology 5, pp. 291-302.
Hall, E.T. (1959): The Silent Language, Doubleday, Garden City, N.Y.
Hall, E.T. (1976): Beyond Culture, Anchor Press, Garden City, N.Y.
Hampton, G. (1979): Students as Subjects in International Behavioral Studies, Journal of International Business Studies 10 [2], pp. 94-96.
Hann, I.; Lee, T.S.; Hui, K. and Png, I.P.L. (2002): Online Information Privacy: Measuring the Cost-Benefit Trade-Off, 23rd International Conference on Information Systems (ICIS), Barcelona, Spain. URL: http://www.comp.nus.edu.sg/~ipng/research/privacy_icis.pdf
Hardin, C. and Banaji, M. (1993): The Influence of Language on Thought, Social Cognition 11, pp. 277-308.
Harris, M.M.; Van Hoye, G. and Lievens, F. (2003): Privacy and Attitudes towards Internet-Based Selection Systems: A Cross-Cultural Comparison, International Journal of Selection and Assessment 11 [2&3], pp. 230-236.
Harris, R.G. (1998): The Economics of Language in a Virtually Integrated Global Economy, Breton, A., Economic Approaches to Language and Bilingualism pp. 35-83, Canadian Heritage, Ottawa.
Hart, W.B. (1996): A Brief History of Intercultural Communication: A Paradigmatic Approach, Speech Communication Association Conference, San Diego, CA, USA. URL: http://web.odu.edu/webroot/inst/AL/wbhart.nsf/pages/histICC

Häubl, G. and Trifts, V. (2000): Consumer Decision Making in Online Shopping Environments: the Effects of Interactive Decision Aids, Marketing Science 19 [1], pp. 4-21.
Hayes, J.R. (1970): Cognition and the Development of Language, Wiley, New York.
Heider, E. (=Rosch) (1971): "Focal" Color Areas and the Development of Color Names, Developmental Psychology 4, pp. 447-455.
Heilmann, K.M. and Scholes, R.J. (1976): The Nature of Comprehension Errors in Broca's Conduction and Wernicke's Aphasics, Cortex 12 [3], pp. 258-265.
Hermeking, M. (2004): Cultural Influence on Internet Diffusion and Website Acceptance, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.
Herring, S. and Estrada, Z. (2004): Representations of Indigenous Language Groups of North and South America on the World Wide Web: In Whose Voice?, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.
Hessler, R.M. and Freerks, K. (1995): Privacy Ethics in the Age of Disclosure: Sweden and America Compared., The American Sociologist 26, pp. 35-53.
Higgins, E.T. (1996): Knowledge Activation: Accessability, Applicability and Salience, Higgins, E.T. and Kruglanski, A.W., Social Psychology: Handbook of Basic Principles pp. 133-168, Guilford, New York.
Hill, T.; Smith, N.D. and Mann, M.F. (1987): Role of Efficacy Expectations in Predicting the Decision to Use Advanced Technologies: the Case of Computers, Journal of Applied Psychology 72, pp. 307-313.
Hjemslev, L. (1943): Prolegomena to a Theory of Language, Waverly Press, Baltimore.
Hoeckling, L. (1995): Managing Cultural Differences: Strategies for Competitive Advantage, Economist Intelligence Unit/ Addison Westley, London.
Hoepfl, M.C. (1997): Choosing Qualitative Research: A Primer for Technology Education Research, Journal of Technology Education 9 [1].
Hoffmann, D.L. and Novak, T.P. (1996): Marketing in Hypermedia Computer Mediated Environments: Conceptual Foundations, Journal of Marketing 60, pp. 50-68.
Hofstede, G. (1980): Culture's Consequences: International Differences in Work-Related Values, Sage, Newbury Park, CA.
Hofstede, G. (1991): Cultures and Organizations: Software of the Mind, McGraw-Hill, London.
Hofstede, G. (2001): Culture's Consequences, 2nd. ed., Sage, Beverly Hills.
Hofstede, G. and McCrae, R. (2004): Personality and Culture Revisited: Linking Traits and Dimensions of Culture, Cross-Cultural Research 2004 [1], pp. 52-88.
Hoft, N. (1996): Developing a Cultural Model, del Galdo, E.M. and Nielsen, J., International User Interfaces pp. 41-73, John Wiley & Sons, New York.
Hogg, M.A.; Joyce, N. and Abrams, D. (1984): Diglossia in Switzerland? A Social Identity Analysis of Speaker Evaluations, Journal of Language and Social Psychology 3 [3], pp. 185-196.
Holden, N.J. (2002): Cross-Cultural Management. A Knowledge Management Perspective, Prentice Hall, Harlow.
Hölscher, C. and Strube, G. (2000): Web Search Behavior of Internet Experts and Newbies, Computer Networks 33 [1-6], pp. 337-346.
Hölscher, U.; Liu, L.; Gruchmann, T.; Pantiskas, C. and Wilcox, S. (2005): Design of Medical Devices for International Markets, Seventh International Workshop on Internationalisation of Products and Systems (IWIPS), Amsterdam, The Netherlands.
Hong, YY.; Morris, MW.; Chin, CY and Benet-Martinez, V. (2000): Multicultural Minds. A Dynamic Constructivistic Approach to Culture and Cognition, American Psychologist 55 [7], pp. 709-720.
Honold, P. (2000): Culture and Context: An Empirical Study for the Development of a Framework for the Elicitation of Cultural Influence in Product Usage, International Journal of Human-Computer Interaction 12 [3&4], pp. 327-345.

Hui, H. and Triandis, H. (1985): Measurement in Cross-Cultural Psychology, Journal of Cross-Cultural Psychology 16 [2], pp. 131-152.
Hunt, E. and Agnoli, F. (1991): The Whorfian Hypothesis: A Cognitive Psychology Perspective, Psychological Review 98 [3], pp. 377-389.
Hyde, K.F. (2000): Recognising Deductive Processes in Qualitative Research, Qualitative Market Research: an International Journal 3 [2], pp. 82-90.
Ingwersen, P. (1982): Search Procedures in the Library Analysed from the Cognitive Point of View, Journal of Documentation 38 [3], pp. 165-191.
Ishikawa, K. (2000): Health Data Use and Protection Policy: Based on Differences by Cultural and Social Environment, International Journal of Medical Informatics 60 [2], pp. 119-125.
Jackson, M.H. (1997): Assessing the Structure of Communication on the World Wide Web, Journal of Computer-Mediated Communication 3 [1], pp. [online], Available at: http://jcmc.indiana.edu/vol3/issue1/jackson.html [Retrieved: October 23, 2005]. URL: http://www.ascusc.org/jcmc/vol3/issue1/jackson.html.
Jacoby, J. and Kaplan, L.B. (1972): The Components of Perceived Risk, 3rd Annual Conference of the Association for Consumer Research, Chicago.
Jagne, J.; Smith, S.G.; Duncker, E. and Curzon, P. (2004): Cross-Cultural Interface Design Strategy, Interaction Design Centre, Middlesex University, Technical Report, IDC-TR-2004-006, http://www.cs.mdx.ac.uk/research/idc/papers/IDC-TR-2004-006.pdf
Jarvenpaa, S.L. and Tractinsky, N. (1999): Consumer Trust in an Internet Store: A Cross-Cultural Validation, Journal of Computer-Mediated Communication 5 [2], pp. [online], Available at: http://jcmc.indiana.edu/vol5/issue2/jarvenpaa.html [Retrieved: November 11, 2005]. URL: http://www.ascusc.org/jcmc/vol5/issue2/jarvenpaa.html
Johnson, C.P.D. (1998): Towards a Meaningful Definition of "Written Culture", The Culture of Publishing [online], pp. Available at: http://apm.brookes.ac.uk/publishing/culture/johnson.html [Retrieved: Juli 6, 2005]. URL: http://apm.brookes.ac.uk/publishing/culture/johnson.html
Kamppuri, M. and Tukiainen, M. (2004): Culture in Human-Computere Interaction Studies: A Survey of Ideas, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.
Karahanna, E.; Evaristo, R. and Srite, M. (2002): Methodological Issues in MIS Cross-Cultural Research, Journal of Global Information Management 10 [1], pp. 48 - 55.
Karahanna, E.; Evaristo, R. and Srite, M. (2005): Cross-Cultural Research in MIS: Issues and Opportunities, Khosrow-Pour, M., Encyclopedia of Information Science and Technology I-V pp. 640-644, Idea Group, Inc., Hershey, PA.
Kashima, E. and Kashima, Y. (1998): The Case of Cultural Dimensions and Personal Pronoun Use, Journal of Cross-Cultural Psychology 29, pp. 461-486.
Keegan, T.; Lewis, R.; Roa, T. and Tarnowska, Joanna (2004): Indigenious Language in an E-Learning Interface: Translation of PLACE into Maori Language, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.
Kiestra, M.D.; Stokmans, M.J.W. and Kamphuis, J. (1994): End-users Searching the Online Catalogue: the Influence of Domain and System Knowledge on Search Patterns, The Electronic Library 12 [6], pp. 335-343.
Kim, K.H.S.; Relking, N.R.; Lee, K.M. and Hirsch, J. (1997): Disctinct Cortical Areas Associated with Native and Second Languages, Nature 338, pp. 171-174.
Kirah, A. (2005): When Culture Meets Technology and When Technology Meets Culture, Seventh International Workshop on Internationalisation of Products and Systems (IWIPS), Amsterdam, The Netherlands.

Kirtley, J.E. (1999): Is Implementing the EU Data Protection Directive in the United States Irreconcilable with the First Amendment?, Government of Information Quarterly 16, pp. 87-92.
Klein, W. and Perdue, C. (1992): Utterance Structure: Developing Grammars Again, John Benjamins, Amsterdam.
Kluckhohn, C.K. (1951): Values and Value Orientation in the Theory of Action, Parsons, T. and Shils, E.A., Toward a General Theory of Action pp. 388-433, Harvard University Press, Cambridge, MA.
Kluckhohn, F.R. and Strodtbeck, F. (1969): Variations in Value Orientations, Row, Peterson, Evanston, Ill.
Ko, H.; Jung, J.; Kim, J. and Shim, S.W. (2004): Cross-Cultural Differences in Perceived Risk of Online Shopping, Journal of Interactive Advertising 4 [2].
Kobsa, A. and Teltzrow, M. (2005): Impacts of Contextualized Communication of Privacy Practices and Personalization Benefits on Purchase Behavior and Perceived Quality of Recommendation, Van Setten, M.; McNean, S. and Konstan, J., Beyond Personalization 2005: A Workshop on the Next Stage of Recommender Systems Research (IUI 2005) pp. 48-53, San Diego, CA, USA.
Koch, T. (2003): Specification for Resource Description Methods: The Role of Classification Schemes in Internet Resource Description and Discovery, 2003, December 23, April 23, 2003, http://www.lub.se/desire/radar/reports/D3.2.3/class_v10.html
Komlodi, A. (2005): Cross-Cultural Study of Information Seeking, 11th International Conference on Human-Computer Interaction (HCI), Las Vegas, NE, USA.
Komlodi, A. and Carlin, M. (2004): Identifying Cultural Variables in Information-Seeking Behavior, Tenth Americas Conference on Information Systems (AMCIS), New York, NY, USA.
Kralisch, A. (2003): Business-Language-Strategien im Internet, Kattenbusch, D., Ed, Mosaicum Scientiae, Haus des Buches, Regensburg.
Kralisch, A. (2005): Information Exchange in Multilingual and Multicultural Virtual Communities, Seventh International Workshop on Internationalisation of Products and Systems (IWIPS), Amsterdam, The Netherlands.
Kralisch, A. and Berendt, B. (2004): Cultural Determinants of Search Behaviour on Websites, Sixth International Workshop on Internationalisation of Products and Systems (IWIPS), Vancouver, Canada.
Kralisch, A. and Berendt, B. (2004): Linguistic Determinants of Search Behaviour on Websites, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.
Kralisch, A. and Berendt, B. (2005): Access to Medical Information in a Cross-linguistic and Cross-cultural Context, General Online Research (G.O.R. 05), Zurich, Switzerland.
Kralisch, A. and Berendt, B. (2005): Language-sensitive Search Behaviour and the Role of Domain Knowledge, New Review of Multimedia and Hypermedia: Special Issue on Minority Language, Multimedia and the Web 11 [2], pp. 221-246.
Kralisch, A. and Eisend, M. (2005): Risk Reduction and Website Satisfaction in a Cross-cultural Context, Internet Research 6.0: Internet Generations (AoIR 2005), Chicago, IL, USA.
Kralisch, A.; Eisend, M. and Berendt, B. (2005): The Impact of Culture on Website Navigation Behaviour, 11th International Conference on Human-Computer Interaction (HCI), Las Vegas, NE, USA.
Kralisch, A. and Köppen, V. (2005): The Impact of Language on Website Use and User Satisfaction: Project Description, 13th European Conference on Information Systems (ECIS), Regensburg, Germany.
Kralisch, A. and Mandl, T. (2006): Barriers to Information Access across Languages on the Internet: Network and Language Effects, 39th Hawaii International Conference on System Sciences (HICSS-39), Poipu, HI, USA.

Kralisch, A. and Mandl, Thomas (2005): Intercultural Aspects of Design and Interaction with Retrieval Systems, 11th International Conference on Human-Computer Interaction (HCI), Las Vegas, NE, USA.
Kralisch, A.; Yeo, A.W. and Jali, N. (2006): Linguistic and Cultural Differences in Information Categorization and their Impact on Website Use, 39th Hawaii International Conference on System Sciences (HICSS-39), Poipu, HI, USA.
Krouglov, A. (2001): Ukrainian: Undoing Previous Corpus and Status Planning, Lingua Planado Kaj Leksikologio, pp. 199-216.
Kuhlthau, C.C. (1993): Seeking Meaning: a Process Approach to Library and Information Services, Ablex, Norwood, NJ.
Lakoff, G. (1987): Woman, Fire and Dangerous Things: What Categories Reveal about the Mind, The University of Chicago Press, Chicago, London.
Lambert, W.E. (1974): Culture and Language as Factors in Learning and Education, Cultural Factors in Learning and Education: Fifth Western Washington Symposium on Learning, Bellingham, WA, USA.
Lamnek, S. (1988): Qualitative Sozialforschung, Band 1: Methodologie, Weihnheim, Munich.
Larson, K. and Czerwinski, M. (1998): Web Page Design: Implications of Memory, Structure and Scent from Information Retrieval, ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '98), Los Angeles, CA, USA.
Laufer, B. (1998): The Development of Passive and Active Vocabulary in a Second Language. Same or Different?, Applied Linguistics 19, pp. 255-271.
Levine, R. and Norenzayan, A. (1999): The Pace of Life in 31 Countries, Journal of Cross-Cultural Psychology 30 [2], pp. 178-205.
Lévi-Strauss, C. (1966): The Savage Mind, University of Chicago Press, Chicago.
Lienert, G.A. (1989): Testaufbau und Testanalyse, VerlagsUnion, München.
Light, A. (2003): Culture and Usability: Will Design Patterns Ease Problems of Context?, Usability News, 2005, November 2, http://www.usabilitynews.com/news/article1185.asp, July 17, 2003, http://www.usabilitynews.com/new/article1185.asp
Lisle, l.; Dong, J. and Isensee, S. (1998): Case Study of Development of an Ease of Use Web Site, Fourth Conference on Human Factors and the Web (HFWeb), Basking Ridge, NJ, USA.
Ljung, M. (1986): Hur ser det ut? Undersökningen in Engelskan in Sveriga, Språkvård 3, pp. 5-10.
Luna, D.; Peracchio, L.A. and de Juan, M.D. (2002): Cross-Cultural and Cognitive Aspects of Web Site Navigation, Journal of the Academy of Marketing Science 30 [4], pp. 397-410.
Luna, D.; Peracchio, L.A. and de Juan, M.D. (2003): The Impact of Language and Congruity on Persuason on Multicultural E-Marketing, Journal of Consumer Psychology 13 [1&2], pp. 41-50.
Luria, A.R. and Vygotsky, L.S. (1992): Ape, Primitive Man, and Child: Essays in the History of Behavior, LEA, Hillsdale.
Lyons, J. (1981): Language and Linguistics: An Introduction, Cambridge University, Cambridge, England.
Mandl, T. (2005): Interkulturelle Aspekte bei der Interaktion mit Informationstechnologie, Beneke, J. and Jarman, F., Interkulturalität in Wissenschaft und Praxis, 15th. ed., pp. 181-190, Universitätsverlag Hildesheim, Hildesheim.
Mandl, T. and Womser-Hacker, C. (2004): Challenges for the Evaluation of Multilingual Information Retrieval Systems, Sixth International Workshop on Internationalisation of Products and Systems (IWIPS), Vancouver, Canada.
Marchionini, G. (1995): Information Seeking in Electronic Environment, Cambridge University Press, Cambridge.
Marchionini, G.; Lin, X. and Dwiggins, S. (1990): Effects of the User's Knowledge on Searching in a Hypertext Environment, 53rd ASIS (American Society for Information Science) Annual Meeting, Toronto, Canada.

Marcus, A. (2004): User-Interface Design and Culture, Aykin, N., Usability and Internationalization of Information Technology pp. 51-78, Lawrence Erlbaum Publishers, New York.
Marcus, A. and West Gould, E. (2000): Crosscurrents. Cultural Dimensions and Global Web User-Interface Design, Interactions 7&8, pp. 32-46.
Marcus, A. and West Gould, E. (2000): Cultural Dimensions and Global Web User-Interface Design: What? So what? Now what?, Sixth Conference on Human Factors and the Web (HFWeb), Austin, TX, USA. URL: http://www.tri.sbc.com/hfweb/marcus/hfweb00_marcus.html
Martinet, André (1962): A Functional View of Language, Oxford University Press, London.
Martins, B. and Marió, S. (2005): Language Identification in Web Pages, ACM Symposium on Applied Computing (SAC), Santa Fe, NM, USA.
Matsumoto, D. (2000): Culture and Psychology: People around the World, Wadsworth/Thompson Learning, Belmont, CA.
McDaniel, M.A.; Hines, R.J. and Guynn, M.J. (2002): When Text Difficulty Benefits Less-Skilled Readers, Journal of Memory and Language 46, pp. 544-561.
McEneany, J.E. (2001): Graphic and Numerical Methods to Assess Navigation in Hypertext, International Journal of Human-Computer Studies 55, pp. 761-786.
McPherson, M.; Smith-Lovin, L. and Cook, J.M. (2001): Birds of a Feather: Homophily in Social Networks, Annual Review of Sociology 27, pp. 415-444.
McSweeney, B. (2002): Hofstede's Model of National Cultural Differences and Consequences: A Triumph of Faith - A Failure of Analysis, Human Relations 55 [1], pp. 89-118.
Meara, P. (1996): The Dimensions of Lexical Competence, Brown, G.; Malmkjaer, K. and William, J., Performance and Competence in Second Language Acquisition pp. 35-53, Cambridge University Press, Cambridge.
Metcalfe, B. (1995): Metcalfe's Law: A Network Becomes More Valuable as it Reaches More Users, Infoworld 17, p. 53.
Mikami, S. and Suzuki, I. (2004): The Language Observatory Project and its Experiment: Cyber Census Survey, Conference on Sharing Capability in Localisation & Human Language Technologies (SCALLA): Crossing the Digital Divide, Kathmandu, Nepal. URL: http://www.elda.org/en/proj/scalla/SCALLA2004/mikami.pdf
Milberg, S.J.; Smith, H.J. and Burke, S.J. (2000): Information Privacy: Corporate Management and National Regulation, Organizational Science 11, pp. 35-37.
Missana, S. The Grip of Culture: Edward T. Hall, 2005, August 29, http://www.ishkbooks.com/hall.pdf
Moe, W.W. (2003): Buying, Searching, or Browsing: Differentiating between Online Shoppers Using In-Store Navigational Clickstream, Journal of Consumer Psychology 13 [1&2], pp. 29-39.
Morden, T. (1999): Models of National Culture - A Management Review, Cross-Cultural Management 6 [1], pp. 19-44.
Mullen, M. (1995): Diagnosing Measurement Equivalence in Cross-National Research, Journal of International Business Studies 3, pp. 573-596.
Nation, P. (1990): Teaching and Learning Vocabulary, Heinle & Heinle, Boston.
Navarro-Prieto, R.; Scaife, M. and Y., Rogers. (1999): Cognitive Strategies in Web Searching, Fifth Conference on Human Factors and the Web (HFWeb), Gaithersburg, Maryland, USA.
Neerincx, M.A.; Lindenberg, J. and Permberton, S. (2001): Support Concepts for Web Navigation: Cognitive Engineering Approach, Tenth World Wide Web Conference (WWW 2001), Hong Kong.
Nielsen, J. (1993): Usability Engineering, Academic Press, Boston.
Nielsen, J. (1997): Search and You May Find, 2005, September 10, July 15, 1997, http://www.useit.com/alertbox/9797b.html
Nielsen, J. (2000): Designing for Web Usabililty, New Riders Publications, Indiana.

Noorderhaven, N.; van Stel, A.; Thurik, R. and Wennekers, S. (1976-2000): Uncertainty Avoidance and the Rate of Business Ownership across 22 OECD Countries, Jena, Max-Planck-Gesellschaft, 13, Discussion Papers on Entrepreneurship, Growth and Public Policy, 1505, ftp://papers.mpiew-jena.mpg.de/egp/discussionpapers/2005-15.pdf

Norman, D. (1983): Some Observations on Mental Models, Gentner, D. and Stevens, A., Mental Models pp. 7-14, Lawrence Erlbaum Associates, Hillsdale, NJ.

Norton, D. (2001): Implementation of an Electronic Report Viewing Application for Multi-Cultural Users, Second South African Human Computer Interaction Conference (CHI-SA), Pretoria, South Africa. URL: South African Human Computer Interaction

Odlyzko, A. and Tilly, B. (2005): A Refutation of Metcalfe's Law and a Better Estimate for the Value of Networks and Network Interconnections, Minneapolis, Digital Technology Center, University of Minnesota, November 2, 2005, http://www.dtc.umn.edu/~odlyzko/doc/metcalfe.pdf

Olsen, C.L. (1979): On Choosing a Test Statistic in Multivariate Analysis of Variance, Psychological Bulletin 83, pp. 579-586.

Otter, M. and Johnson, M. (2000): Lost in Hyperspace: Metrics and Mentals Models, Interacting with Computers 13 [1], pp. 1-40.

Palmer, J.W.; Bailey, J.P. and Faraj, S. (2000): The Role of Intermediaries in the Development of Trust on the WWW: The Use and Prominence of Trusted Third Parties and Privacy Statements, Journal of Computer-Mediated Communication 5 [3], pp. [online], Available at: http://www.ascusc.org/jcmc/vol5/issue3/palmer.htm [Retrieved: November 1, 2005].

Paolillo, J. (1999): The Virtual Speech Community: Social Network and Language Variation on IRC, Journal of Computer-Mediated Communication 4 [4], p. [online] Available at: http://jcmc.indiana.edu/vol4/issue4/paolillo.html. URL: http://jcmc.indiana.edu/vol4/issue4/paolillo.html

Pargman, D. and Palme, J. (2004): Linguistic Standardization on the Internet, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.

Park, H.W. (2002): E-Science and Hyperlink Network Analysis: Collaborative Communication through Hyperlinking, Conference of the Netherlands School of Communication Research, Utrecht, The Netherlands.

Park, H.W. (2003): Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web, Connections 25 [1], pp. 49-61. URL: http://www.insna.org/Connections-Web/Volume25-1/7.Hyperlink.pdf

Park, H.W. and Thelwall, M. (2003): Hyperlink Analyses of the World Wide Web: A Review, Journal of Computer-Mediated Communication 8 [4], pp. [online], Available at: http://jcmc.indiana.edu/vol8/issue4/park.html [Retrieved: July 17, 2003].

Perugini, S.; McDevitt, K.; Richardson, R.; Perez-Quinones, M.; Shen, R.; Ramakrishnan, N.; Williams, C. and Fox, E.A. (2004): Enhancing Usability in CITIDEL: Multimodal, Multilingual, and Interactive Visualization Interfaces, Joint Conference on Digital Libraries (JCDL), Tuscon, AZ, USA.

Peters, E. and Slovic, P. (1996): The Role of Affect and Worldview as Orienting Dispositions in the Perception and Acceptance of Nuclear Power, Journal of Applied Social Psychology 26, pp. 1427-1453.

Petty, R.E. and Cacioppo, J.T. (1986): Communications and Persuasion: Central and Peripheral Routes to Attitude Change, Springer-Verlag, New York.

Pinkerton, B. (1994): Finding what People Want: Experiences with the WebCrawler, First International World-Wide Web Conference, Geneva, Switzerland.

Pirolli, P. and Card, S. (1995): Information Foraging in Information Access Environments, ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'95), Denver, Colorado, USA.

Pirolli, P. and Card, S. (1999): Information Foraging, Psychological Review 106 [4], pp. 643-675.
Preece, J.; Rogers, Y.; Sharp, H.; Benyon, D.; Holland, S. and Carey, T. (1994): Human-Computer Interaction, Addison-Westley, Harlow, England.
Punj, G. and Staelin, R. (1983): A Model of Consumer Information Search Behaviour for New Automobiles, Journal of Consumer Research 9, pp. 366-379.
Reeder, K.; Macfadyen, L.P.; Chase, M. and Roche, J. (2004): Falling through the (Cultural) Gaps? Intercultural Communication Challenges in Cyberspace, Fourth International Conference on Cultural Attitudes towards Technology and Communication (CATaC), Karlstad, Sweden.
Reips, U.-D. (1999): Theorie und Techniken des Web-Experimentierens, Batinic, B.; Gräf, L.; Werner, A. and Bandilla, W., Online Research: Methoden, Anwendungen und Ergebnisse pp. 277-295, Hofgrefe, Göttingen.
Rice, R.E. and Barnett, G. (1986): Group Communication Networking in an Informatioon Environment: Applying Metrics Multidimensional Scaling, McLaughlin, M., Communication Yearbook 9, Sage, Beverly Hills, CA.
Rice, R.E.; McCreadie, M. and Chang, S.-J.L. (2001): Accessing and Browsing Information and Communication, MIT Press, Cambridge, MA.
Rokeach, M. (1979): Understanding Human Values: Individual and Societal, The Free Press, New York.
Romberg, M.; Röse, K. and Zühlke, D. (1999): Global Demands of Non-European Markets for the Design of User-Interfaces, MMI-Interaktiv 1, pp. [online], Available at: www.doaj.org/ftxt/useworld.net/ausgaben/3-1999/romberg.pdf [Retrieved: September 28, 2005]. URL: http://useworld.net/ausgaben/3-1999/romberg.pdf
Rosch, E. (1975): Cognitive Reference Points, Cognitive Psychology 7, pp. 532-547.
Röse, K. (2005): Usability Engineering in a German Context - How Cultural Values Can Influence Product Design, 11th International Conference on Human-Computer Interaction (HCI), Las Vegas, NE, USA.
Rosenberg, M.J. and Hovland, C.I. (1960): Cognitive, Affective, and Behavioural Components of Attitudes, Hovland, C.I. and Rosenberg, M.J., Attitude Organization and Change, Yale University Press, New Haven.
Russo, P. and Boor, S. (1993): How Fluent is your Interface? Designing for International Users, Conference on Human Factors in Computing Systems (INTERCHI), Amsterdam, The Netherlands.
Rustemli, A. and Kokdemir, D. (1993): Privacy Dimensions and Preferences among Turkish Students, Journal of Social Psychology 133, pp. 807-814.
Sandra, D. and Rice, S. (1995): Network Analyses of Prepositional Meaning: Mirroring whose Mind - the Linguist's or the Language User's?, Cognitive Linguistics 6 [1], pp. 89-130.
Sapir, E. (1921): Language: An Introduction to the Study of Speech, Harcourt, Brace and Company, New York.
Sapir, E. (1933): Language, Encyclopaedia of the Social Sciences, New York.
Sapir, E. (1958): Culture, Language, and Personality, Mandelbaum, D.G., Ed, University of California Press, Berkely, CA.
Sarris, V. (1992): Methodologische Grundlagen der Experimentalpsychologie, Reinhardt, München.
Schein, E. (1985): Organizational Culture and Leadership, Jossey-Bass, San Francisco, CA.
Schur, E. (2002): Density and Complexity in the Semantic Networks of Native and Non-Native Speakers, Second Language Vocabulary Acquisition Colloquium, Leiden, The Netherlands.
Schwartz, S.H. (1992): Universals in the Content and Structure of Value: Theoretical Advances and Empirical Tests in 20 Countries, Zonna, M.P., Advances in Experimental Social Psychology pp. 1-65, Academic Press, San Diego.
Shadish, W.R.; Cook, T.D. and Campbell, D.T. (2002): Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Houghton Mifflin, Boston, MA.

Shahabi, C.; Zarkesh, A.M.; Adibi, J. and Shah, V. (1997): Knowledge Discovery from Users Web-Page Navigation, Seventh International Workship on Research Issues in Data Engineering (IEEE RIDE), Birmingham, UK.
Shannon, C.E. and Weaver, W. (1949): A Mathematical Model of Communication, University of Illinois Press, Urbana, UL.
Shaw, M. and Wright, J. (1967): Scales for the Measurement of Attitudes, McGraw-Hill, New York.
Shneiderman, B. (1997): Designing Information-Abundant Web Sites: Issues and Recommendations, International Journal of Human-Computer Studies 47 [1], pp. 5-30.
Sims, J.; Powell, P. and Vidgen, R. (2005): E-Learning and the Digital Divide: Perpetuating Cultural and Socio-Economic Elitism in Higher Education, 13th European Conference on Information Systems (ECIS), Regensburg, Germany.
Slay, J. and Burke, M.M. (2001): A World View Analysis of the Effect of Culture within C2, Sixth International Command and Control Research and Technology Conference, Annapolis, MD, USA. URL: www.unisa.edu.au/seec/pubs/01papers/051_tr7-slay-burke.pdf
Smith, H.J. (2001): Information Privacy and Marketing: What the US Should (and Shouldn't) Learn from Europe, California Management Review 43, pp. 8-33.
Smith, P.A. (1996): Towards a Practical Measure of Hypertext Usability, Interacting with Computers 8 [4], pp. 365-381.
Smith, R.; Olsen, J.A. and Harris, A. (1999): Resource Allocation Decisions and the Use of Willingness-to-pay as a Valuation Technique within Economic Evaluation: Recommendations from a Review of the Literature, Monas University: Centre for Health Program Evaluation, http://www.buseco.monash.edu.au/centres/che/pubs/wp87.pdf
Snowberry, K.; Parkinson, S. and Sission, N. (1983): Computer Display Menus, Ergonomics 26 [7], pp. 699-712.
Spencer-Oatey, H. (2000): Culturally Speaking: Managing Rapport through Talk across Cultures, Continuum, London.
Spiliopoulou, M.; Bamshad, M.; Berendt, B. and Nakagawa, M. (2003): A Framework for the Evaluation of Session Reconstruction Heuristics in Web-Usage Analysis, INFORMS Journal on Computing 15 [2], pp. 171-190.
Spilioupoulou, M. and Pohle, C. (2001): Data Mining for Measuring and Improving the Success of Web Sites, Journal of Data Mining and Knowledge Discovery 5 [1-2], pp. 85-114.
Spolsky, B. (1989): Conditions for Second Language Learning: Introduction to a General Theory, Oxford University Press, Oxford.
Srinivasan, N. and Rathford, B.T. (1991): An Empirical Test of a Model of External Search for Automobiles, Journal of Consumer Research 18, pp. 233-242.
Srite, M.; Straub, D.; Loch, K.; Evaristo, J.R. and Karahanna, E. (2003): Inquiry into Definitions of Culture in IT Studies, Tan, F., Advanced Topics in Global Information Management pp. 30-48, Idea Group Publishing, Hershey, PA.
Stander, A. and van Belle, J.-P. (2004): Attitudes towards and the Need for Multi-Lingual Websites in South Africa, Cultural Attitudes towards Technology and Communication (CATaC 2004), Karlstad, Sweden.
Stangl, W. (1989): Das neue Paradigma der Psychologie. Die Psychologie im Diskurs des Radikalen Konstruktivismus., Vieweg, Braunschweig.
Steffenson, M.S.; Joag-Dev, C. and Anderson, R.C. (1979): A Cross-Cultural Perspective on Reading Comprehension, Reading Research Quarterly 15 [1], pp. 10-29.
Stone, E.F. and Stone, D.L. (1990): Privacy in Organizations: Theoretical Issues, Research Findings, and Protection Mechanisms, Research in Personnel and Human Resources Management 8, pp. 349-411.
Sun, H. (2003): Cultural Usability: A Localisation Study of Mobile Text Messaging Use, ACM SIGCHI Conference on Human Factors in Computer Systems (CHI '03), Extended Abstracts, Fort Lauderdale, FL, USA.

Sundaram, D.S. and Ronald, T. (1998): An Investigation of External Information Search Effort: Replication in In-home Shopping Situations, Advances in Consumer Research 25 [1], pp. 440-445.
Sunstein, C. (2001): Republic.com, Princeton University Press, Princeton.
Swan, R. and Allan, J. (1998): Aspect Windows, 3-D Visualisation, and Indirect Comparisons of Information Retrieval Systems, 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Melbourne, Australia.
Symons, S. and Pressley, M. (1993): Prior Knowledge Affects Text Search Success and Extraction of Information, Reading Research Quarterly 28, pp. 250-259.
Tajfel, H. (1978): Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations, Academic Press, London.
Tam, J.C. (2000): Personal Data Privacy in the Asia Pacific: A Real Possibility, Tenth International Conference on Computers, Freedom, and Privacy, Toronto, Canada. URL: www.cfp2000.org/papers/tam.pdf
Tan, P.N. and Kumar, V. (2002): Discovery of Web Robot Sessions Based on their Navigational Patterns, Data Mining and Knowledge Discovery 6, pp. 9-35.
Tanenhaus, M.K. and Trueswell, J.C. (1995): Sentence Comprehension, Miller, J. and Eimas, P., Handbook of Perception and Cognition: Speech, Language and Cognition pp. 217-262, Academic Press, San Diego.
Taylor, C.R.; Franke, G.R. and Maynard, M.L. (2001): Attitudes towards Direct Marketing and its Regulation: A Comparison of the United States and Japan, Journal of Public Policy and Marketing 19, pp. 228-237.
Taylor, J.R. (1995): Linguistic Categorisation: Prototypes in Linguistic Theory, Oxford University Press, Oxford.
Taylor, R. (1962): The Process of Asking Questions, American Documentation 13 [4], pp. 391-397.
Terveen, L. and Hill, W. (1998): Evaluating Emergent Collaboration on the Web, Conference of Computer Supported Cooperative Work (CSCW), Seattle, WA, USA.
Trochim, W.M. (2000): Research Methods Knowledge Base - Survey Research, 2005, August 29, May 12, 2000, http://trochim.human.cornell.edu/kb/survey.htm,
Troman, G.; Jeffrey, B. and Walford, G. (2005): Methodological Issues and Practices in Ethnography, Studies in Educational Ethnography 11, JAI Press, Walford.
Trompenaars, F. and Hampden Turner, C. (1993): Riding the Waves of Culture, The Economist Press, London.
Tylor, E.B. (1871): Primitive Culture, Encyclopedia Britannica.
Vakkari, P. (2000): ECognition and Changes of Search Terms and Tactics during Task Performance: A Longitudinal Case Study, Sixth International Conference on Computer-Assisted Information Retrieval (RIAO 2000), Paris, France.
Van De Vijver, J.R. (2000): Methodological Issues in Psychological Research on Culture, Journal of Cross-Cultural Psychology 31 [1], pp. 33-51.
Van Dyke Parunak, H. (1989): Hypermedia Topologies and User Navigation, Hypertext '89, Pittsburgh, PA, USA.
Warschauer, M. (1998): Technology and Indigenous Language Revitalization. Analyzing the Experience of Hawai'i, Canadian Modern Language Review 55 [1], pp. 140-161.
Warschauer, M.; El Said, G.R. and Zohry, A. (2002): Language Choice Online: Globalization and Identity in Egypt, Journal of Computer-Mediated Communication 7 [4], pp. [online], Available at: http://jcmc.indiana.edu/vol7/issue4/warschauer.html [Retrieved: November 1, 2005]. URL: http://jcmc.indiana.edu/vol7/issue4/warschauer.html
Webb, E.J.; Campbell, D.T.; Schwart, R.D.; Sechrest, L. and Grove, J.B. (1981): Nonreactive Measures in the Social Sciences, Houghton Mifflin, Boston.

Weber, E.U.; Hsee, C.K. and Sokolowska, J. (1998): What Fokore Tells Us about Risk and Risk Taking: Cross-Cultural Comparisons of American, German, and Chinese Proverbs, Organizational Behavior and Human Decision Processes 75 [2], pp. 170-186.
Webley, P. and Nyhus, E. (1999): Representation of Savings of Savings Behaviour, 2005, October 27, August 27, 1999, http://www.ex.ac.uk/~PWebley/papers/saving.html
White, D. and Korotayev, A. (2004): Statistics for Cross-Cultural Research, 2005, December 14, http://eclectic.ss.uci.edu/~drwhite/xc/!XC-BK1.doc - http://eclectic.ss.uci.edu/~drwhite/xc/!XC-BK11.doc
Whorf, B.L. (1956): Language, Thought and Reality, Carroll, J.B., Ed, MIT Press, Cambridge, MA.
Williamson, D. (2002): Forward from a Critique of Hofstede's Model of National Culture, Human Relations 55 [1], pp. 1373-1395.
Wong, R.Y and Hong, Y. (2005): Influences of Culture on Cooperation in the Prisoner's Dilemma, American Psychological Society 6 [6], pp. 429-434.
Wright, S. (2000): Community and Communication: The Role of Language in Nation State Building and European Integration, Multilingual Matters Ltd., Clevedon.
Yee, I.H. (1993): Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers, Journal of the American Society for Information Science 44 [3], pp. 161-174.
Yeo, A. and Loo, W. (2004): Identification and Evaluation of Classification Schemes: A User-Centred Approach, Sixth International Workshop on Internationalisation of Products and Systems (IWIPS), Vancouver, Canada.
Yoo, Y.; Ginzberg, M.J. and Ahn, J.H. (1999): A Cross-Cultural Investigation of the Use of Knowledge Management Systems, 20th International Conference on Information Systems (ICIS), Charlotte, NC, USA.
Zook, M.A. (2001): Old Hierarchies of New Networks of Centrality? - The Global Geography of the Internet Content Market, American Behavioral Scientist 44 [10], pp. 1679-1696.

7. APPENDICES

Chapter 1

A-1.1 Brief Desciption of Selected Cultural Dimensions

According to Hofstede and summarized by Baumgartner and Marcus for the purpose of cross-cultural IS website design research (Baumgartner, 2003; Marcus and West Gould, 2000)

Hofstede

INDIVIDUALISM & COLLECTIVISM (IND)

The dimension of Individualism vs. collectivism refers to the role of the individual and group, and which interest prevails over the other (Baumgartner, 2003). Individualism in culture implies loose ties; everyone is expected to look after one's self or immediate family but no one else. Collectivism implies that people are integrated from birth into strong, cohesive groups that protect them in exchange for unquestioned loyalty. Individualistic societies and governments place individual social-economic-interests over the group, maintain strong rights to privacy, nurture strong private opinions, and restrain the power of the state in the economy (Marcus and West Gould, 2000).

UNCERTAINTY AVOIDANCE (UA)

Hofstede's dimension of Uncertainty Avoidance can be defined as the extent to which the members of a culture feel threatened by uncertain or unknown situations (Baumgartner, 2003; Hofstede, 1991). Cultures vary in their avoidance of uncertainty, creating different rituals and having different values regarding formality, punctuality, legal-religious-social requirements, and tolerance for ambiguity (Marcus and West Gould, 2000; Marcus and West Gould, 2000).

POWER DISTANCE (PD)

Power Distance focuses on the nature of human relationship in terms of hierarchy. It describes the extent to which less powerful members of institutions and organizations accept that power is distributed unequally. (Baumgartner, 2003; Hofstede, 1980; Hofstede, 2001). Hofstede claims that high PD countries tend to have centralized political power and exhibit tall hierarchies in

organizations with large differences in salary and status. Low PD countries tend to view subordinates and supervisors as closer together and more interchangeable, with flatter hierarchies in organizations and less difference in salaries and status (Marcus and West Gould, 2000).

MASCULINITY (MAS)

Hofstede describes Masculinity as pertaining to societies in which social gender roles are clearly distinct (i.e., men are supposed to be assertive, tough, and focussed on material success, whereas women are supposed to be more modest, tender, and concerned with the quality of life). Feminity pertains to societies in which gender role overlap (i.e., both men and women are supposed to be modest, tender, and concerned with the quality of life) (Baumgartner, 2003).

LONG-TERM VS. SHORT-TERM ORIENTATION (LTO)

Long-term Orientation stands for the fostering of virtues oriented towards the future rewards, in particular perseverance and thrift. Its opposite pole, Short Term Orientation, stands for the fostering of virtues related to the past and present, in particular, respects for tradition, preservation of 'face' and fulfilling social obligations. (Baumgartner, 2003; Hofstede, 1980; 2001).

Hall

CONTEXT SPECIFICITY

Context refers to the amount and specificity of information in a given situation. A high context communication is one in which most of the meaning is in the context while very little is the transmitted message. A low context communication is similar interacting with a computer – if the information is not explicitely statd; and the program followed religiously, the meaning is distorted (Baumgartner, 2003; Hall, 1976).

MONOCHRONIC VS. POLYCHRONIC

Monochronic (sequential) cultures are characterized by the isolation of activities. People tend to do one thing at a time. They view time as a narrow line of distinct, consecutive segments. … They strongly prefer planning and keeping to plans once they have been made. The opposite approach towards structuring time is synchronism (polychronic cultures). People structuring time synchronically usually do several things at a time, and plans are easily changed (Hall, 1959).

A-1.2 Cultural Index Scores

Indexes from: Hofstede (1991). *Cultures and Organizations: Software of the Mind.* London: McGraw-Hill.

PDI=Power Distance Index Score
IND=Individualism Index Score
MAS = Masculinity Index Score
UAI=Uncertainty Avoidance Index Score
LTO=Long-term Orientation Index Score

	PDI		IND		MAS		UAI		LTO	
	rank	score	rank	score	rank	score	rank	score	rank	Score
Arab Countries	7	80	26/27	38	23	53	27	68		
Argentina	35/36	49	22/23	46	20/21	56	10/15	86		
Australia	41	36	2	90	16	61	37	51	15	31
Austria	53	11	18	55	2	79	24/25	70		
Bangladesh									11	40
Belgium	20	65	8	75	22	54	5/6	94		
Brazil	14	69	26/27	38	27	49	21/22	76	6	65
Canada	39	39	4/5	80	24	52	41/42	48	20	23
Chile	24/25	63	38	23	46	28	10/15	86		
China									1	118
Columbia	17	67	49	13	11/12	64	20	80		
Costa Rica	42/44	35	46	15	48/49	21	10/15	86		
Denmark	51	18	9	74	50	16	51	23		
East Africa	21/23	64	33/35	27	39	41	36	52		
Equador	8/9	78	52	8	13/14	63	28	67		
Finland	46	33	17	63	47	26	31/32	59		
France	15/16	68	10/11	71	35/36	43	10/15	86		
Germany FR	42/44	35	15	67	9/10	66	29	65	14	31
Great Britain	42/44	35	3	89	9/10	66	47/48	35	18	25
Greece	27/28	60	30	35	18/19	57	1	112		
Guatemala	2/3	95	53	6	43	37	3	101		
Hong Kong	15/16	68	37	25	18/19	57	49/50	29	2	96
India	10/11	77	21	48	20/21	56	45	40	7	61
Indonesia	8/9	78	47/48	14	30/31	46	41/42	48		

Iran	29/30	58	24	41	35/36	43	31/32	59		
Ireland (Rep of)	49	28	12	70	7/8	68	47/48	35		
Israel	52	13	19	54	29	47	19	81		
Italy	34	50	7	76	4/5	70	23	75		
Jamaica	37	45	25	39	7/8	68	52	13		
Japan	33	54	22/23	46	1	95	7	92	4	80
Malaysia	1	104	36	26	25/26	50	46	36		
Mexico	5/6	81	32	30	6	69	18	82		
Netherlands	40	38	4/5	80	51	14	35	53	10	44
New Zealand	50	22	6	79	17	58	39/40	49	16	30
Nigeria									22	16
Norway	47/48	31	13	69	52	8	38	50		
Pakistan	32	55	47/48	14	25/26	50	24/25	70	23	0
Panama	2/3	95	51	11	34	44	10/15	86		
Peru	21/23	64	45	16	37/38	42	9	87		
Philippines	4	94	31	32	11/12	64	44	44	21	19
Poland									13	32
Portugal	24/25	63	33/35	27	45	31	2	104		
Salvador	18/19	66	42	19	40	40	5/6	94		
Singapore	13	74	39/41	20	28	48	53	8	9	48
South Africa	35/36	49	16	65	13/14	63	39/40	49		
South Korea	27/28	60	43	18	41	39	16/17	85	5	75
Spain	31	57	20	51	37/38	42	10/15	86		
Sweden	47/48	31	10/11	71	53	5	49/50	29	12	33
Switzerland	45	34	14	68	4/5	70	33	58		
Taiwan	29/30	58	44	17	32/33	45	26	69	3	87
Thailand	21/23	64	39/41	20	44	34	30	64	8	56
Turkey	18/19	66	28	37	32/3	45	16/17	85		
Uruguay	26	61	29	36	42	38	4	100		
USA	38	40	1	91	15	62	43	46	17	29
Venezuela	5/6	81	50	12	3	73	21/22	76		
West Africa	10/11	77	39/41	20	30/31	46	34	54		
Yugoslavia	12	76	33/35	27	48/49	21	8	88		
Zimbabwe									19	25

A-2.1 Example of MINT-Query for WUM

```
select t
from node as a b, template b * a as t
where a.url contains "XYZ"
and b.url contains "XYZ"
```

“XYZ” replaces the chosen URL-sections.

A-2.2 Use of Search Options: Results from Cultural Group Comparisons

	UA		*Context*		*LTO*		*PD*	
	Low	*High*	*Low*	*High*	*Low*	*High*	*Low*	*High*
Measure 1: % of users who used the…								
Search engine	29.1%	30.5%**	29.8%	21.0%**	29.7%	23.8%**	29.8%	28.0%**
Alphabetical search	55.1%	58.8%**	55.5%	55.2% (p=0.875)	55.8%	52.7%**	56.1%	57.3%*
Content search	48.3%	45.0%**	47.0%	49.7% (p=0.075)	47.1%	53.6%**	46.9%	48.9%**
Measure 2: Absolute frequency of use of …								
Search engine	0.51 (1.307)	**0.55**** (1.298)	**0.51** (1.287)	0.39* (1.312)	0.52 (1.291)	0.50 (1.469)	0.52 (1.300)	0.53 (1.325)
Alphabetical search	2.04 (4.020)	2.42** (4.456)	2.03 (3.957)	1.61** (2.990)	2.05 (3.966)	2.18 (4.624)	2.11 (4.065)	**2.46**** (4.673)
Content search	0.72 (1.125)	0.73 (1.342)	0.71 (1.147)	0.76 (1.201)	0.71 (1.128)	**0.85**** (1.229)	0.72 (1.187)	**0.78**** (1.273)
Measure 3: Relative frequency of use of …								
Search enginea)	1/14.01 (16.138)	1/14.07 (16.510)	1/13.78 (15.92)	**1/10.87*** (11.73)	1/13.90 (16.17)	1/13.04 (14.938)	1/14.03 (16.340)	1/14.02 (15.882)
Alphabetical search	0.32 (0.293)	0.27 (0.226)	0.33 (0.299)	0.281 (0.231)	0.32 (0.290)	0.25 (0.205)	**0.31** (0.284)	0.25* (0.203)
Content search	0.19 (0.237)	0.17 (0.203)	0.19 (0.243)	**0.24**** (0.253)	**0.31** (0.288)	0.23** (0.244)	0.18 (0.230)	**0.19*** (0.211)

(only users who used at least one search option were considered)

Grey : no hypothesis about the distribution

For reasons of better comparibility, numbers are provided in the reciprocal form.

** significant at the 0.001 level; * significant at the 0.05 level

The MANOVA results for the first measure (search option used or not) were: $F(1,83260)=17.681$, $p<0.001$ (UA, search engine); $F(1,83260)=104.582$, $p<0.001$ (UA, alphabetical search); $F(1,83260)=82.690$, $p<0.001$ (UA, content search); $F(1,57688)=39.311$, $p=0.000$ (Context, search engine); $F(1,57688)=0.025$, $p<0.875$ (Context, alphabetical search): $F(1,57688)=3.179$, $p=0.075$ (Context, content search); $F(1, 67265)=59.862$, $p<0.001$ (LTO, search engine); $F(1, 67265)=13.736$, $p<0.001$ (LTO, alphabetical search); $F(1, 67265)=60.829$, $p<0.001$ (LTO, content search); $F(1, 83260)=18.173$, $p<0.001$ (PD, search engine); $F(1, 83260)=6.762$, $p=0.009$ (PD, alphabetical search); $F(1, 83260)=19.307$, $p<0.001$ (PD, content search);

For the absolute frequency we obtained: $F(1,83260)=15.844$, $p<0.001$ (UA, search engine); $F(1,83260)=157.842$, $p<0.001$ (UA, alphabetical search); $F(1,83260)=0.695$, $p<0.405$ (UA, content search); $F(1,57688)=9.739$, $p=0.002$ (Context, search engine);$F(1,57688)=12.260$, $p<0.001$ (Context, alphabetical search): $F(1,57688)=1.732$, $p=0.188$ (Context, content search); $F(1, 67265)=0.686$, $p=0.407$ (LTO, search engine); $F(1, 67265)=3.793$, $p=0.051$ (LTO, alphabetical search); $F(1, 67265)=54.964$, $p<0.001$ (LTO, content search); $F(1, 83260)=0.578$, $p=0.447$ (PD, search engine); $F(1, 83260)=81.281$, $p<0.001$ (PD, alphabetical search); $F(1, 83260)=35.771$, $p<0.001$ (PD, content search).

The relative frequency has the following values: $F(1,11451)=0.143$, $p=0.705$ (UA, search engine); $F(1,11451)=1.426$, $p=0.232$ (UA, alphabetical search); $F(1,11451)=0.258$, $p=0.611$ (UA, content search); $F(1,7658)=4.187$, $p=0.041$ (Context, search engine);$F(1,7658)=0.616$, $p<0.433$ (Context, alphabetical search): $F(1,7658)=13.574$, $p<0.001$ (Context, content search); $F(1, 2.091) =2.091$, $p=0.148$ (LTO, search engine); $F(1, 8995)=0.054$, $p=0.816$ (LTO, alphabetical search); $F(1, 8995)=8.184$, $p=0.004$ (LTO, content search); $F(1, 11451)=2.707$, $p=0.100$ (PD, search engine); $F(1, 11451)=4.119$, $p=0.042$ (PD, alphabetical search); $F(1, 11451)=3.890$, $p=0.049$ (PD, content search).

A-2.3 Questionnaire: Preference for Information Categorization (English Version)

Task 1

We would like you to arrange the cards lying on your desk into several groups. Write the numbers of the pictures (see the back side of the pictures) that belong to the same group into the same space, separated by commas (,). You are free to choose the number of groups and subgroups and the number of cards per group. The order of the groups is not important.

Notice: Try not to spend much more than 10 minutes for this task.

Groups	Picture Id Insertion (back side of your cards)
Group 1:	
Group 2:	
Group 3:	
Group 4:	
Group 5:	
Group 6:	

Task 2: Group

Grouped pictures were listed here.

Please give the groups intuitive names. Can you give reasons for your chosen way of grouping the cards? What are the characteristics of each group and subgroup? Please describe with a few words.

Name of Group:

I put them into one group because...

Task 3

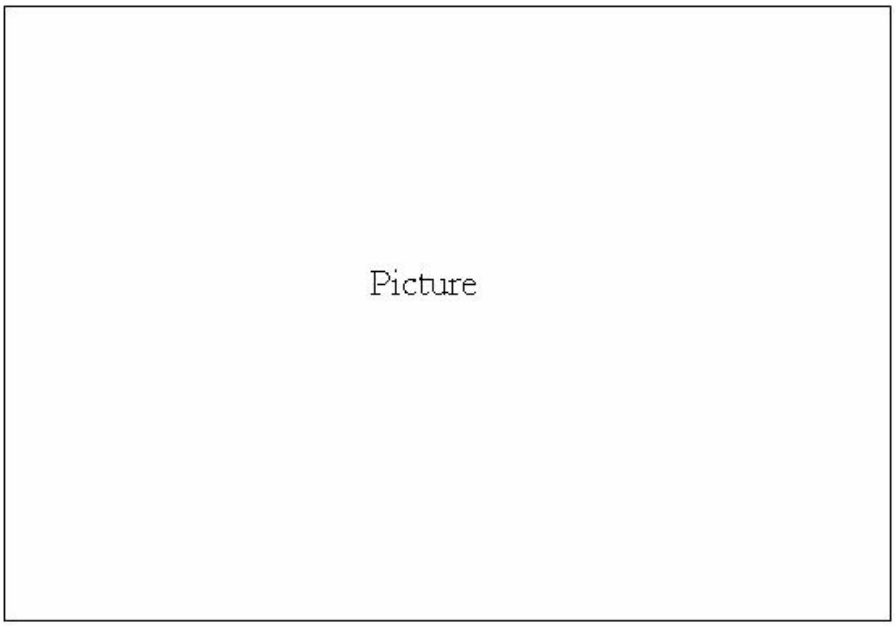
Imagine you had the following search categories on a website about dermatological diseases. Please rank the following categorization cues according to your preferences for using them. Write down the number 1 in front of the categorization cue with the highest preference, 2 in front of the next, etc.

1. 1 Alphabetical Search
2. 1 Search of body part where the disease occurs
3. 1 Colour of disease symptoms
4. 1 Severity of disease symptoms
5. 1 Non-dermatological symptoms accompanying the disease (e.g. fever)
6. 1 Causes for disease (e.g. bacterial infection, genetically inherited, etc.)
7. 1 Dermatological symptoms of disease (rash, callus, etc.)

Submit

Please answer the following questions using the website www.siteA.com and your language version. **Use right mouse button popup menu to open this site in a new window.** (The German language version for German and Russian participants. The English language version for British, Irish, and Malaysian participants.)

If the task takes you more than 10 minutes, please write as answer "I couldn't find the answer."



What is the name of the disease?

On which body parts does it predominantly occur?

Is this disease considered to be a differential diagnosis of Lichen spinulosus?

Submit

Task 4a

Why did you choose (please select) as the first search option for your previous search?

- I usually start with this search option.
- I was hoping to get an overview with that search option.
- I thought it would be the fastest way for finding the results.
- I thought it would be the easiest way for finding the results.
- I saw it first.
- It happened by accident.

Submit

Please indicate on the following scale to which extent you agree with th following statements.
Note 1: You have used a search option once you have clicked on it.
Note 2: Please base your judgement on the previous search task and the given time-limit.

---

a) Alphabetical Search. ☐ I did not use the alphabetical search. (Please continue with b).)

	I strongly agree					I don't agree at all
	1	2	3	4	5	6
Using the alphabetical search permitted me to find the *correct* answers easily.	○	○	○	○	○	○
The alphabetical search enabled me to accomplish the tasks fast.	○	○	○	○	○	○
The alphabetical search allowed me to find the answers in an efficient way.	○	○	○	○	○	○
The alphabetical search allowed me to accomplish the tasks faster than it probably would have been with other search options.	○	○	○	○	○	○
Overall, I found the alphabetical search useful for accomplishing the tasks.	○	○	○	○	○	○

---

b) Search by localization. ☐ I did not search by localization. (Please continue with c).)

	I strongly agree					I don't agree at all
	1	2	3	4	5	6
Using the search-by-localization tool permitted me to find the correct answers efficiently.	○	○	○	○	○	○
The search-by-localization tool enabled me to accomplish the tasks fast.	○	○	○	○	○	○
The search-by-localization tool allowed me to find the answers in an efficient way.	○	○	○	○	○	○
The search-by-localization tool allowed me to accomplish the tasks faster than it probably would have been with other search options.	○	○	○	○	○	○
Overall, I found the search-by-localization tool useful for accomplishing the tasks.	○	○	○	○	○	○

---

c) Search engine. ☐ I did not use the search engine. (Please continue with d).)

	I strongly agree					I don't agree at all
	1	2	3	4	5	6
Using the search engine permitted me to find the *correct* answers easily.	○	○	○	○	○	○
The search engine enabled me to accomplish the tasks fast.	○	○	○	○	○	○
The search engine allowed me to find the answers in an efficient way.	○	○	○	○	○	○
The search engine allowed me to accomplish the tasks faster than it probably would have been with other search options.	○	○	○	○	○	○
Overall, I found the search engine useful for accomplishing the tasks.	○	○	○	○	○	○

d) All search options

	I strongly agree					I don't agree at all
	1	2	3	4	5	6
Every search option permitted me to find the *correct* answers easily.	○	○	○	○	○	○
All search options enabled me to accomplish the tasks fast.	○	○	○	○	○	○
All search options allowed me to find the answers in an efficient way.	○	○	○	○	○	○
Overall, I found all search options useful for accomplishing the tasks.	○	○	○	○	○	○

Submit

Please indicate on the following scale to which extent you agree with the following statements.

---

a) Alphabetical Search ☐ I didn't use the alphabetical search. Please continue with b)

	I strongly agree					I don't agree at all
	1	2	3	4	5	6
I find the alphabetic search cumbersome to use.	○	○	○	○	○	○
It took me only a short moment to get along with the alphabetical search.	○	○	○	○	○	○
It was frustrating for me to use the alphabetical search.	○	○	○	○	○	○
Overall, I found it easy to use the alphabetical search.	○	○	○	○	○	○

b) Search by localization ☐ I didn't search by localization. (If so, please continue with c).)

	I strongly agree					I don't agree at all
	1	2	3	4	5	6
I find the search-by-localization tool cumbersome to use.	○	○	○	○	○	○
It took me only a short moment to get along with the search-by-localization tool.	○	○	○	○	○	○
It was frustrating for me to use the search-by-localization tool.	○	○	○	○	○	○
Overall, I found it easy to use the search-by-localization tool.	○	○	○	○	○	○

c) Search engine ☐ I didn't use the search engine. (If so, please continue with d).)

	I strongly agree					I don't agree at all
	1	2	3	4	5	6
I find the search engine cumbersome to use.	○	○	○	○	○	○
It took me only a short moment to get along with the search engine.	○	○	○	○	○	○
It was frustrating for me to use the search engine.	○	○	○	○	○	○
Overall, I found it easy to use the search engine.	○	○	○	○	○	○

d) All Search Options

	I strongly agree					I don't agree at all
	1	2	3	4	5	6
I find all search options cumbersome to use.	○	○	○	○	○	○
It took me only a short moment to get along with all of the search options.	○	○	○	○	○	○
It was frustrating for me to use any of the provided search options.	○	○	○	○	○	○
Overall, I found it easy to use any of the provided search options.	○	○	○	○	○	○

Submit

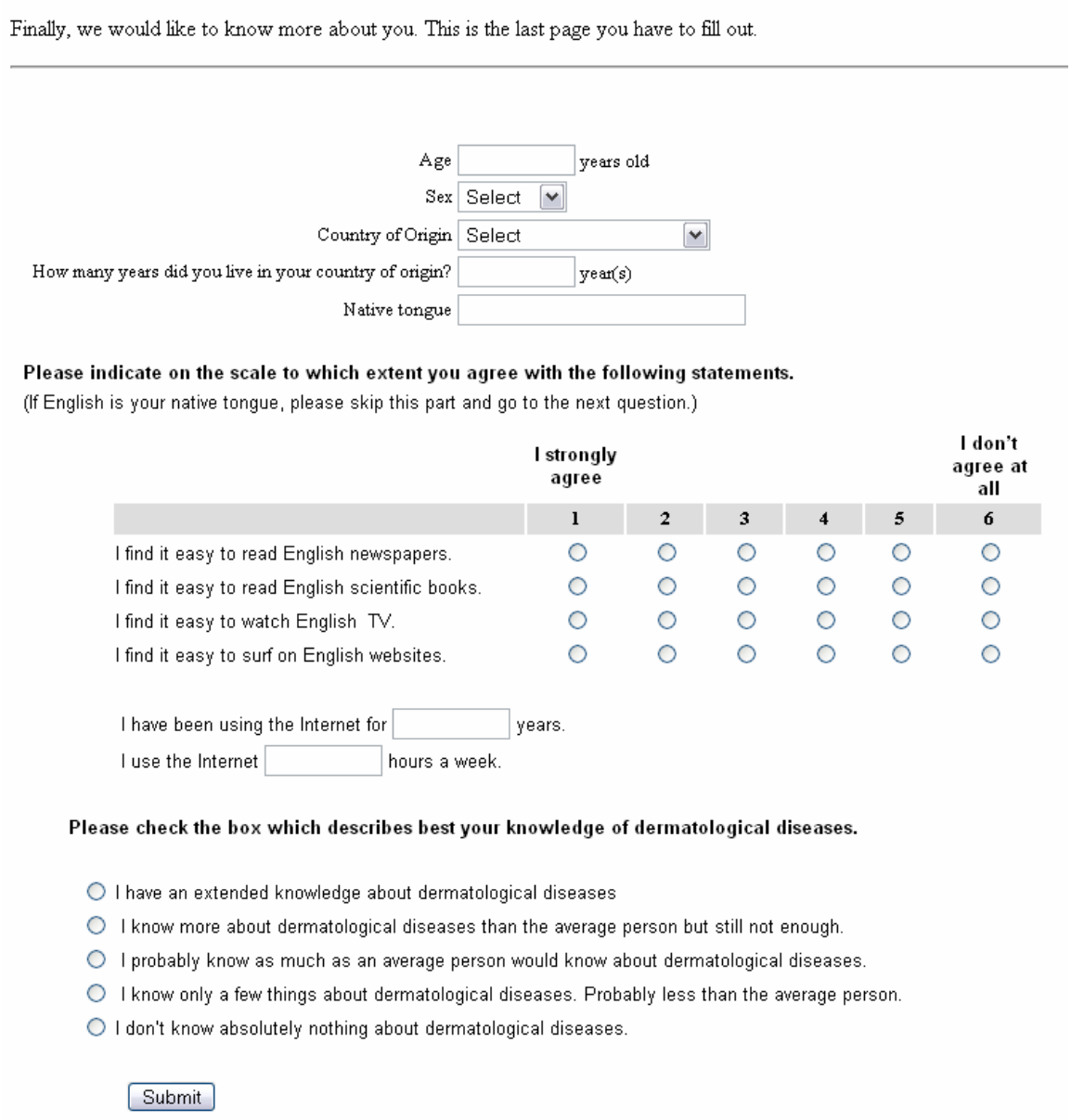
All things considered, for finding information about a dermatological disease using a **search tool that classifies by dermatological symptoms** (e.g. rash, callus) would be:

	extremely	quite	slightly	slightly	quite	extremely	
bad	○	○	○	○	○	○	**good**
foolish	○	○	○	○	○	○	**wise**
unfavorable	○	○	○	○	○	○	**favorable**
harmful	○	○	○	○	○	○	**beneficial**
negative	○	○	○	○	○	○	**positive**

All things considered, for finding information about a dermatological disease using **a search tool that classifies by colours of the symptoms** would be:

	extremely	quite	slightly	slightly	quite	extremely	
bad	○	○	○	○	○	○	**good**
foolish	○	○	○	○	○	○	**wise**
unfavorable	○	○	○	○	○	○	**favorable**
harmful	○	○	○	○	○	○	**beneficial**
negative	○	○	○	○	○	○	**positive**

Submit



Finally, we would like to know more about you. This is the last page you have to fill out.

Age ____ years old
Sex Select
Country of Origin Select
How many years did you live in your country of origin? ____ year(s)
Native tongue ____

Please indicate on the scale to which extent you agree with the following statements.
(If English is your native tongue, please skip this part and go to the next question.)

	I strongly agree 1	2	3	4	5	I don't agree at all 6
I find it easy to read English newspapers.	○	○	○	○	○	○
I find it easy to read English scientific books.	○	○	○	○	○	○
I find it easy to watch English TV.	○	○	○	○	○	○
I find it easy to surf on English websites.	○	○	○	○	○	○

I have been using the Internet for ____ years.
I use the Internet ____ hours a week.

Please check the box which describes best your knowledge of dermatological diseases.

○ I have an extended knowledge about dermatological diseases
○ I know more about dermatological diseases than the average person but still not enough.
○ I probably know as much as an average person would know about dermatological diseases.
○ I know only a few things about dermatological diseases. Probably less than the average person.
○ I don't know absolutely nothing about dermatological diseases.

Submit

A-2.4 Selected Statistics: Preferences for Information Categorization (Study 4)

(1) Which criteria were used?

Group	*Symptom*	*Name*	*Body part*	*Cause*	*Colour*	*Severity*
1 (L1/G)	31.8%	13.4%	30.2%	8.9%	30.7%	5%
3 (L2/G)	24.2%	7.5%	38.3%	18.3%	17.5%	8.3%
2 (L1/E)	10.0%	0%	84.3%	4.3%	27.1%	2.9%

(2) Kruskal-Wallis-Test: Ranking Order

Ranks

	Group	*N*	*Mean Rank*
rank_alph	1 (L1/G)	31	29,23
	3 (L2/G)	15	24,00
	2 (L1/E)	9	30,44
	Total	55	
rank_body	1 (L1/G)	31	28,24
	3 (L2/G)	15	29,97
	2 (L1/E)	9	23,89
	Total	55	
rank_colour	1 (L1/G)	31	26,76
	3 (L2/G)	15	32,60
	2 (L1/E)	9	24,61
	Total	55	
rank_severity	1 (L1/G)	31	29,95
	3 (L2/G)	15	27,43
	2 (L1/E)	9	22,22
	Total	55	
rank_accsymp	1 (L1/G)	31	26,48
	3 (L2/G)	15	30,40
	2 (L1/E)	9	29,22
	Total	55	
rank_cause	1 (L1/G)	31	26,76
	3 (L2/G)	15	26,67
	2 (L1/E)	9	34,50
	Total	55	
rank_dermsymp	1 (L1/G)	31	27,40
	3 (L2/G)	15	30,77
	2 (L1/E)	9	25,44
	Total	55	

Test Statistics(a,b)

	rank_alph	*rank_body*	*rank_colour*	*rank_severity*	*rank_accsymp*	*rank_cause*	*rank_dermsymp*
X^2	1,474	,933	1,948	1,764	,698	1,834	,806
Df	2	2	2	2	2	2	2
Asymp. Sig.	,479	,627	,377	,414	,706	,400	,668

a Kruskal Wallis Test

b Grouping Variable: mother tongue in numbers

(3) Kruskal-Wallis-Test: Use of Search Criteria: **Ranks**

	Group	*N*	*Mean Rank*
symptom is used as criterion	1 (L1/G)	179	1355,82
	3 (L2/G)	232	1276,13
	2 (L1/E)	210	1101,45
	4 (L2/E)	1708	1137,72
	Total	2329	
Name is used as criterion	1 (L1/G)	179	1262,13
	3 (L2/G)	232	1196,35
	2 (L1/E)	210	1106,00
	4 (L2/E)	1708	1157,82
	Total	2329	
Body part is used as criterion	1 (L1/G)	179	892,80
	3 (L2/G)	232	1003,28
	2 (L1/E)	210	1523,01
	4 (L2/E)	1708	1171,48
	Total	2329	
Cause is used as criterion	1 (L1/G)	179	1176,59
	3 (L2/G)	232	1293,35
	2 (L1/E)	210	1122,41
	4 (L2/E)	1708	1151,59
	Total	2329	
colour is used as criterion	1 (L1/G)	179	1315,81
	3 (L2/G)	232	1168,81
	2 (L1/E)	210	1274,08
	4 (L2/E)	1708	1135,27
	Total	2329	

Test Statistics(a,b)

	symptom is used as criterion	*name is used as criterion*	*body part is used as criterion*	*cause is used as criterion*	*colour is used as criterion*
Chi-Square	64,876	41,930	137,252	45,708	40,764
Df	3	3	3	3	3
Asymp. Sig.	,000	,000	,000	,000	,000

a Kruskal Wallis Test

b Grouping Variable: group

(3a) Use of Search Criteria: Group 1 (L1/G) vs. Group 3 (L2/G)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
symptom is used as criterion	Equal variances assumed	9,013	,003	1,533	409	,126	,068	,045	-,019	,156
	Equal variances not assumed			1,519	368,154	,130	,068	,045	-,020	,157
name is used as criterion	Equal variances assumed	14,266	,000	1,878	409	,061	,056	,030	-,003	,116
	Equal variances not assumed			1,821	329,901	,069	,056	,031	-,005	,118
body part is used as criterion	Equal variances assumed	16,307	,000	-1,998	409	,046	-,095	,047	-,188	-,002
	Equal variances not assumed			-2,014	393,607	,045	-,095	,047	-,187	-,002
cause is used as criterion	Equal variances assumed	36,662	,000	-2,876	409	,004	-,100	,035	-,169	-,032
	Equal variances not assumed			-2,993	407,712	,003	-,100	,034	-,166	-,034
colour is used as criterion	Equal variances assumed	35,080	,000	3,014	409	,003	,126	,042	,044	,209
	Equal variances not assumed			2,945	343,993	,003	,126	,043	,042	,211

(3b) Use of Search Criteria: Group 1 (L1/G) vs. Group 2 (L1/E)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
symptom is used as criterion	Equal variances assumed	140,196	,000	5,559	387	,000	,218	,039	,141	,296
	Equal variances not assumed			5,378	294,633	,000	,218	,041	,138	,298
name is used as criterion	Equal variances assumed	181,151	,000	5,688	387	,000	,134	,024	,088	,180
	Equal variances not assumed			5,250	178,000	,000	,134	,026	,084	,184
body part is used as criterion	Equal variances assumed	48,140	,000	-12,929	387	,000	-,541	,042	-,623	-,459
	Equal variances not assumed			-12,695	337,295	,000	-,541	,043	-,625	-,457
cause is used as criterion	Equal variances assumed	14,321	,000	1,869	387	,062	,047	,025	-,002	,095
	Equal variances not assumed			1,820	314,281	,070	,047	,026	-,004	,097
colour is used as criterion	Equal variances assumed	2,378	,124	,777	387	,438	,036	,046	-,055	,127
	Equal variances not assumed			,774	372,505	,439	,036	,046	-,055	,127

(3c) Use of Search Criteria: Group 1 (L1/G) vs. Group 4 (L2/E)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
symptom is used as criterion	Equal variances assumed	116,915	,000	6,774	1885	,000	,187	,028	,133	,242
	Equal variances not assumed			5,223	197,962	,000	,187	,036	,117	,258
name is used as criterion	Equal variances assumed	92,645	,000	5,123	1885	,000	,090	,017	,055	,124
	Equal variances not assumed			3,442	191,825	,001	,090	,026	,038	,141
body part is used as criterion	Equal variances assumed	197,206	,000	-6,154	1885	,000	-,239	,039	-,316	-,163
	Equal variances not assumed			-6,564	224,094	,000	-,239	,036	-,311	-,167
cause is used as criterion	Equal variances assumed	4,415	,036	1,071	1885	,284	,021	,020	-,018	,061
	Equal variances not assumed			,966	207,899	,335	,021	,022	-,022	,065
colour is used as criterion	Equal variances assumed	76,392	,000	5,329	1885	,000	,155	,029	,098	,212
	Equal variances not assumed			4,348	201,135	,000	,155	,036	,085	,225

(3d) Use of Search Criteria: Group 3 (L2/G) vs. Group 2 (L1/E)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
symptom is used as criterion	Equal variances assumed	80,284	,000	4,182	440	,000	,150	,036	,080	,220
	Equal variances not assumed			4,256	412,727	,000	,150	,035	,081	,219
name is used as criterion	Equal variances assumed	83,846	,000	4,193	440	,000	,078	,019	,041	,114
	Equal variances not assumed			4,408	231,000	,000	,078	,018	,043	,112
body part is used as criterion	Equal variances assumed	143,310	,000	-10,767	440	,000	-,446	,041	-,528	-,365
	Equal variances not assumed			-10,922	424,494	,000	-,446	,041	-,527	-,366
cause is used as criterion	Equal variances assumed	117,615	,000	4,859	440	,000	,147	,030	,087	,206
	Equal variances not assumed			5,001	353,403	,000	,147	,029	,089	,205
colour is used as criterion	Equal variances assumed	20,957	,000	-2,285	440	,023	-,090	,040	-,168	-,013
	Equal variances not assumed			-2,268	415,663	,024	-,090	,040	-,169	-,012

(3e) Use of Search Criteria: Group 3 (L2/G) vs. Group 4 (L2/E)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
symptom is used as criterion	Equal variances assumed	73,345	,000	4,846	1938	,000	,119	,025	,071	,167
	Equal variances not assumed			4,010	270,310	,000	,119	,030	,061	,177
name is used as criterion	Equal variances assumed	18,528	,000	2,204	1938	,028	,033	,015	,004	,063
	Equal variances not assumed			1,809	269,400	,072	,033	,018	-,003	,069
body part is used as criterion	Equal variances assumed	24,661	,000	-4,149	1938	,000	-,144	,035	-,213	-,076
	Equal variances not assumed			-4,202	299,633	,000	-,144	,034	-,212	-,077
cause is used as criterion	Equal variances assumed	136,905	,000	6,388	1938	,000	,122	,019	,084	,159
	Equal variances not assumed			4,593	257,362	,000	,122	,027	,070	,174
colour is used as criterion	Equal variances assumed	4,843	,028	1,135	1938	,256	,029	,025	-,021	,079
	Equal variances not assumed			1,076	288,084	,283	,029	,027	-,024	,082

(3f) Use of Search Criteria: Group 2 (L1/E) vs. Group 4 (L2/E)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
symptom is used as criterion	Equal variances assumed	6,965	,008	-1,276	1916	,202	-,031	,024	-,079	,017
	Equal variances not assumed			-1,397	278,001	,164	-,031	,022	-,075	,013
name is used as criterion	Equal variances assumed	42,987	,000	-3,126	1916	,002	-,044	,014	-,072	-,017
	Equal variances not assumed			-8,916	1707,000	,000	-,044	,005	-,054	-,035
body part is used as criterion	Equal variances assumed	1208,108	,000	8,500	1916	,000	,302	,036	,232	,372
	Equal variances not assumed			10,814	313,939	,000	,302	,028	,247	,357
cause is used as criterion	Equal variances assumed	8,092	,004	-1,388	1916	,165	-,025	,018	-,060	,010
	Equal variances not assumed			-1,640	294,155	,102	-,025	,015	-,055	,005
colour is used as criterion	Equal variances assumed	58,664	,000	4,409	1916	,000	,119	,027	,066	,172
	Equal variances not assumed			3,729	243,542	,000	,119	,032	,056	,182

(4) Ease of Use

Report

factor: ease alphabet

group	Mean	N	Std. Deviation
1	1,2679887	10	,92281821
3	,7184898	12	,81483132
2	1,3642935	6	,77894199
4	-,2730326	108	,84004302
Total	,0000000	136	1,00000000

Low Value = Higher Ease of Use

Report

factor: ease body parts

Group	Mean	N	Std. Deviation
1	,7478902	21	,82712835
3	,0082258	28	1,03854106
2	1,2442822	24	1,09430255
4	-,2602204	176	,81706747
Total	,0000000	249	1,00000000

Report

factor: ease search engine

Group	Mean	N	Std. Deviation
1	,6377264	15	1,33784103
3	1,0017572	14	1,29029872
2	,4125205	15	,95148385
4	-,1731297	172	,86439292
Total	,0000000	216	1,00000000

(5) Usefulness

Ranks

	group	N	Mean Rank
Factor Analysis: Usefulness alphabet	1	16	98,38
	3	10	84,20
	2	9	54,00
	4	172	108,29
	Total	207	
Factor Analysis: Usefulness body part	1	16	147,06
	3	10	118,90
	2	9	62,00
	4	172	101,33
	Total	207	
Factor Analysis: Usefulness search engine	1	16	150,00
	3	10	151,80
	2	9	149,33
	4	172	94,57
	Total	207	
Factor Analysis: Usefulness all	1	16	163,22
	3	10	135,80
	2	9	127,17
	4	172	95,43
	Total	207	

Low Value = Higher Usefulness

(4a) Usefulness: Group 1 (L1/G) vs. Group 3 (L2/G) (Preference for Alphabetical Search)

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Factor Analysis: Usefulness alphabet	3	6	2,1930965	,21308246	,08699055
	1	9	1,0281550	1,46670565	,48890188

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Factor Analysis: Usefulness alphabet	Equal variances assumed	75,530	,000	1,909	13	,079	1,16494149	,61039472	-,15373614	2,48361912
	Equal variances not assumed			2,346	8,501	,045	1,16494149	,49658072	,03146415	2,29841883

(4b) Usefulness: Group 2 (L1/E) vs. Group 3 (L2/G) (Preference for Alphabetical Search)

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Factor Analysis: Usefulness alphabet	3	6	2,1930965	,21308246	,08699055
	2	3	2,4681758	,00000000	,00000000

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Factor Analysis: Usefulness alphabet	Equal variances assumed	18,656	,003	-2,160	7	,068	-,27507937	,12734113	-,57619328	,02603454
	Equal variances not assumed			-3,162	5,000	,025	-,27507937	,08699055	-,49869570	-,05146304

(4c) Usefulness: Group 4 (L2/E) and Other Groups (Preference for Body Part)

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Factor Analysis: Usefulness alphabet	>= 4	312	-,0174227	,83664224	,04736554
	< 4	88	-,0140658	1,17706278	,12547531
Factor Analysis: Usefulness body part	>= 4	312	-,3272392	,64263445	,03638201
	< 4	88	,4465726	1,18851151	,12669575
Factor Analysis: Usefulness search engine	>= 4	312	-,1141147	,68002513	,03849884
	< 4	88	,0378888	1,23378106	,13152150
Factor Analysis: Usefulness all	>= 4	312	-,4974944	,68618431	,03884754
	< 4	88	,7025884	,95538463	,10184434

(4d) Usefulness: Group 1 (L1/G) vs. Group 3 (L2/G) (Preference Body Part Group 3 (L2/G))

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Factor Analysis: Usefulness body part	1	20	1,3172214	,92905866	,20774383
	3	24	,3562114	,59244681	,12093270

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Factor Analysis: Usefulness body part	Equal variances assumed	3,035	,089	4,158	42	,000	,96100999	,23111172	,49460767	1,42741232
	Equal variances not assumed			3,998	31,108	,000	,96100999	,24037932	,47082204	1,45119794

(4e) Usefulness: Group 1 (L1/G) vs. Group 2 (L1/E) (no significant differences in Preferences for Body Part

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Factor Analysis: Usefulness body part	1	20	1,3172214	,92905866	,20774383
	2	18	1,5283108	,76498528	,18030876

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Factor Analysis: Usefulness body part	Equal variances assumed	,662	,421	-,759	36	,453	-,21108937	,27794927	-,77479660	,35261787
	Equal variances not assumed			-,767	35,740	,448	-,21108937	,27507953	-,76911740	,34693867

(5) Attitude

Report

group		*Factor analysis: alphabet*	*Factor analysis: body parts*	*Factor analysis: search engine*	*Factor analysis: symptoms*	*Factor analysis: colour*
1	Mean	-,1193104	,2897316	-,2546937	,3153397	,2248258
	N	32	32	32	32	32
	Std. Deviation	1,00136646	,73946124	,89088306	,94174481	,85454191
3	Mean	-,5447177	,2079412	-,2611166	-,0584885	,1519675
	N	40	40	40	40	40
	Std. Deviation	,79656369	,94500288	,99559642	1,09363501	1,07334775
2	Mean	-,0122503	,1259865	-,0482507	,5553885	,5193273
	N	33	33	33	33	33
	Std. Deviation	1,04782997	,78106691	,94774843	,75530393	,97839081
4	Mean	,1903467	-,1899498	,1782473	-,1926971	-,2044407
	N	312	312	312	312	312
	Std. Deviation	,98616253	1,09125567	1,01780387	,97267438	,99115190
Total	Mean	,0800414	-,0899706	,0849545	-,0816363	-,0800350
	N	417	417	417	417	417
	Std. Deviation	,99805065	1,04521970	1,01190056	,99147098	1,01139164

(5a) Attitude: Group 3 (L2/G) vs. Group 4 (L2/E) (Attitude towards Colour and Alphabetical Search)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Factor analysis: alphabet	Equal variances assumed	3,711	,055	-4,527	350	,000	-,73506432	,16238122	-1,05443002	-,41569863
	Equal variances not assumed			-5,336	55,564	,000	-,73506432	,13776750	-1,01109360	-,45903504
Factor analysis: body parts	Equal variances assumed	1,497	,222	2,202	350	,028	,39789104	,18069808	,04250039	,75328170
	Equal variances not assumed			2,461	53,279	,017	,39789104	,16168659	,07362851	,72215358
Factor analysis: search engine	Equal variances assumed	,104	,747	-2,577	350	,010	-,43936386	,17052233	-,77474121	-,10398651
	Equal variances not assumed			-2,621	50,039	,012	-,43936386	,16763227	-,77605674	-,10267098
Factor analysis: symptoms	Equal variances assumed	,245	,621	,810	350	,419	,13420868	,16574159	-,19176609	,46018345
	Equal variances not assumed			,740	47,250	,463	,13420868	,18147533	-,23082153	,49923889
Factor analysis: colour	Equal variances assumed	,076	,783	2,121	350	,035	,35640824	,16805218	,02588910	,68692738
	Equal variances not assumed			1,994	47,921	,052	,35640824	,17874716	-,00300224	,71581873

(6) Use

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
not_alph_final	2	33	,82	,392	,068
	3	40	,70	,464	,073
not_body_final	2	33	,27	,452	,079
	3	40	,30	,464	,073
not_search_final	2	33	,55	,506	,088
	3	40	,65	,483	,076

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
not_alph_final	1	32	,69	,471	,083
	3	40	,70	,464	,073
not_body_final	1	32	,34	,483	,085
	3	40	,30	,464	,073
not_search_final	1	32	,53	,507	,090
	3	40	,65	,483	,076

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
not_alph_final	1	32	,69	,471	,083
	4	312	,65	,477	,027
not_body_final	1	32	,34	,483	,085
	4	312	,44	,497	,028
not_search_final	1	32	,53	,507	,090
	4	312	,45	,498	,028

(6a) Use: Group 4 (L2/E) vs. Group 3 (L2/G) (Measure 1)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
not_alph_final	Equal variances assumed	1,649	,200	-,578	350	,563	-,046	,080	-,203	,111
	Equal variances not assumed			-,590	50,140	,558	-,046	,078	-,203	,111
not_body_final	Equal variances assumed	24,545	,000	1,641	350	,102	,136	,083	-,027	,299
	Equal variances not assumed			1,729	51,156	,090	,136	,079	-,022	,294
not_search_final	Equal variances assumed	12,017	,001	-2,414	350	,016	-,201	,083	-,365	-,037
	Equal variances not assumed			-2,472	50,244	,017	-,201	,081	-,365	-,038

(6b) Group 4 (L2/E) vs. Group 2 (L1/E) (Measure 1)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
not_alph_final	Equal variances assumed	28,406	,000	-1,913	343	,057	-,164	,086	-,333	,005
	Equal variances not assumed			-2,241	42,696	,030	-,164	,073	-,312	-,016
not_body_final	Equal variances assumed	35,416	,000	1,809	343	,071	,163	,090	-,014	,341
	Equal variances not assumed			1,952	40,616	,058	,163	,084	-,006	,332
not_search_final	Equal variances assumed	,015	,903	-1,059	343	,290	-,097	,091	-,276	,083
	Equal variances not assumed			-1,047	38,865	,302	-,097	,092	-,284	,090

(6b) Use: Measure 2 (first use)

Report

group		*alphabetical search used first*	*se_first*	*body_first*
1	Mean	,13	,19	,59
	N	32	32	32
	Std. Deviation	,336	,397	,499
3	Mean	,05	,20	,65
	N	40	40	40
	Std. Deviation	,221	,405	,483
2	Mean	,18	,27	,55
	N	33	33	33
	Std. Deviation	,392	,452	,506
4	Mean	,10	,32	,47
	N	312	312	312
	Std. Deviation	,304	,467	,500
Total	Mean	,11	,29	,51
	N	417	417	417
	Std. Deviation	,308	,457	,501

A-3.1 Questionnaire: Risk Perception-Information Needs (English Version)

(relevant sections)

2a. Information

Progress

A medical problem may cause various problems for a lot of people.
Assume you could not visit this website: **HOW LIKELY** is it that you would meet the following problems.

Please indicate your answers on a scale from 1 (Rather likely) to 7 (Rather unlikely).

	Rather likely						Rather unlikely
	1	2	3	4	5	6	7
Financial loss (e.g. doctor's fee, treatment)	○	○	○	○	○	○	○
Physical pain	○	○	○	○	○	○	○
Psychological problems (e.g. mental suffering with neuro-dermatitis)	○	○	○	○	○	○	○
Social and interpersonal problemes (e.g. talking about the disease; visible symptoms contacts with others, etc.)	○	○	○	○	○	○	○
Problems in professional life (e.g. lost job prospects, job loss, etc.)	○	○	○	○	○	○	○

2b. Information

Progress

Different types of stress and strain affect people in different ways.
What about you? How **cumbersome** are/would the following aspects be?

	Hardly cumbersome/ hardly relevant						Very cumbersome/ very relevant
	1	2	3	4	5	6	7
Financial loss e.g. doctor's fee, treatment)	○	○	○	○	○	○	○
Physical pain	○	○	○	○	○	○	○
Psychological problems (e.g. mental suffering with neuro-dermatitis)	○	○	○	○	○	○	○
Social and interpersonal problemes (e.g. talking about the disease; visible symptoms contacts with others, etc.)	○	○	○	○	○	○	○
Problems in professional life (e.g. lost job prospects, job loss, etc.)	○	○	○	○	○	○	○

2c. Information

On a scale from 1 (I completely agree.) to 7 (I completely disagree.), to which extent do you agree with the following statements?

	I completely agree.						I disagree completely.
	1	2	3	4	5	6	7
It is worth consulting more than one information source for a medical problem.	○	○	○	○	○	○	○
I am sure to come to a better decision when I gather more information.	○	○	○	○	○	○	○
I know which treatment is the best for me when I extensively collect information.	○	○	○	○	○	○	○
There is much to lose if you don't inform yourself enough.	○	○	○	○	○	○	○

3a. Website

Please rate the information on the website by the means of the following adjectives. For each pair, indicate which adjective best describes the information on this website.

☐ *I couldn't say. I haven't surfed on this website yet.*

Please click on the "Continue"-Button at the end of this page.

You can differentiate your judgement on a scale from 1 to 7.

	1	2	3	4	5	6	7	
reliable	○	○	○	○	○	○	○	unreliable
not trustworthy	○	○	○	○	○	○	○	trustworthy
accurate	○	○	○	○	○	○	○	inaccurate
objective	○	○	○	○	○	○	○	subjective
useful	○	○	○	○	○	○	○	useless
unimportant	○	○	○	○	○	○	○	important
relevant	○	○	○	○	○	○	○	irrelevant
out dated	○	○	○	○	○	○	○	up to date
incomplete	○	○	○	○	○	○	○	complete
incomprehensible	○	○	○	○	○	○	○	comprehensible
easy to access	○	○	○	○	○	○	○	rather difficult to access

3b. Website

Progress

Please rate your **process of searching for information** on this website, by the means of the following adjective pairs.
Indicate, for each pair, which adjective best describes your information search.

You can differentiate your judgement on a scale from 1 to 7.

	1	2	3	4	5	6	7	
efficient	○	○	○	○	○	○	○	inefficient
rather quick	○	○	○	○	○	○	○	rather time-consuming
successful	○	○	○	○	○	○	○	unsuccessful

Please rate **the result achieved through your visit** to this website by the means of the following adjectives.
Indicate, for each pair, which adjective best describes the result of your visit.

You can differentiate your judgement on a scale from 1 to 7.

	1	2	3	4	5	6	7	
unsatisfied	○	○	○	○	○	○	○	satisfied
pleased	○	○	○	○	○	○	○	disappointed
positive	○	○	○	○	○	○	○	negative
discontented	○	○	○	○	○	○	○	contented

4. Demographic Data

Last Page!

Finally, we would like to ask you some general questions.
Please answer the following questions by clicking on the applicable answer from the list.

Where did you grow up? --Please select--

In which country do you presently reside? --Please select--

Your age: (Please type in.)

Your gender: ☐ male ☐ female

Do you work in the field of medicine? --Please select--

How many years have you been using the Internet? for --Please select--

How often have you used the Internet last week? --Please select--

Why do you visit this website? --Please select--

A-3.2 Regression Results: Regression Coefficients (standard errors) and Model Results: Risk Perception – Information Need (Study 5)

	Perceived risk	*Expected benefits of search*	*Search efforts*	*Perceived information quality*	*Perceived search efficiency*	*Perceived search results*
Constant	127.164(30.782)***	4.634(.388)***	219.971(175.528)	4.088(.609)***	-.616(1.220)	-.022(.490)
Individualism vs. collectivism	.237(.213)	-	-	-	-	-
Uncertainty avoidance	.027(.225)	-	-.238(.952)	-	-	.012(.006)*
Masculinity vs. femininity	-.464(.223)*	.014(.005)***	-	-	-	-
Long-term orientation	-	-	-.238(.952)	-	-.003(.013)	-
Perceived risk	-	-	.399(.933)	-	-	-
Expected benefits of search	-	-	-23.803(20.204)	-	-	-
Search efforts	-	-	-	-	.007(.003)*	-
Perceived information quality	-	-	-	-	1.173(.301)***	-
Perceived search efficiency	-	-	-	-	-	.787(.174)***
Quality of health care system	-7.909(3.756)**	-	-	-.166(.109)	-	-
Domain Knowledge	-	-	-	-.009(.004)**	-	-
Web experience	-	.011(.003)***	-	-.009(.004)**	-	-
n[1]	36	37	24	36	23	23
R^2	.230	.312	.048	.294	.465	.698
F	2.32*	7.87***	.92	4.44***	6.36***	18.45***

* significant at p < .1

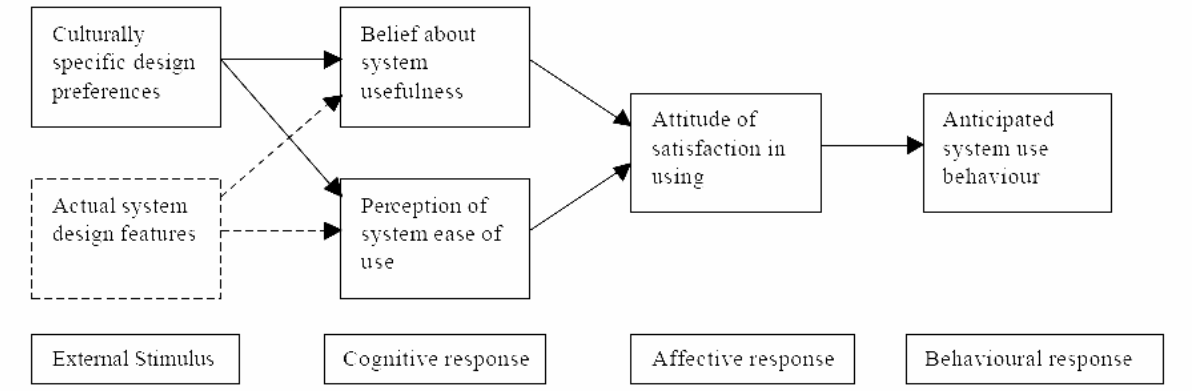
** significant at p < .05

*** significant at p < .01

[1] sample sizes differ since cultural indices for all countries are not available

A-3.3 Adaptation of Davis' TAM Model to a Cross-cultural Background

Evers & Day (Evers and Day, 1997)



A-3.4 Questionnaire (Website Survey): Product Valuation (English Version) (relevant sections)

2. Paying for Online Content

progress

In answering the questions in this section, please refer to your general opinion regarding paying for online content.
Please indicate your agreement/disagreement with each of the following statements.

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
All informational content available on the internet should be free of charge.	○	○	○	○	○
There is certain information available online that should be paid for.	○	○	○	○	○
People should have unlimited and free of charge access to educational content online.	○	○	○	○	○
All health information available on the internet should be free of charge.	○	○	○	○	○
Certain health information available on the internet should be paid for.	○	○	○	○	○
Paying for a service on the internet usually means the service is of higher quality.	○	○	○	○	○

3. Evaluation of www.website.net

The questions in this section deal with your usage experience of website.net **in particular.**

Following are a number of adjectives concerning aspects of the website. For each adjective pair, please indicate on a scale from 1 to 6 which adjective best describes your usage experience.

The information I received from the website...		
uses too few medical terms	○1 ○2 ○3 ○4 ○5 ○6	uses too many medical terms
is well illustrated	○1 ○2 ○3 ○4 ○5 ○6	is poorly illustrated
satisfied my information needs	○1 ○2 ○3 ○4 ○5 ○6	left my information needs unfulfilled
is informative	○1 ○2 ○3 ○4 ○5 ○6	is vague
is conveyed using consumer friendly language	○1 ○2 ○3 ○4 ○5 ○6	is conveyed using consumer unfriendly language
is supported with sufficient hyperlinks to other websites	○1 ○2 ○3 ○4 ○5 ○6	is not supported sufficiently with hyperlinks
is easy to find	○1 ○2 ○3 ○4 ○5 ○6	is hard to find
is supported with sufficient information about alternative treatments	○1 ○2 ○3 ○4 ○5 ○6	is not supported sufficiently with information about alternative treatments
provides standard information	○1 ○2 ○3 ○4 ○5 ○6	provides alternative information
is easy to access	○1 ○2 ○3 ○4 ○5 ○6	hard to access
is adequately supported by pictures	○1 ○2 ○3 ○4 ○5 ○6	is inadequately supported by pictures
is subjective	○1 ○2 ○3 ○4 ○5 ○6	is objective
is interesting	○1 ○2 ○3 ○4 ○5 ○6	is boring
has impressive graphic support	○1 ○2 ○3 ○4 ○5 ○6	has unexciting graphic support
is unique	○1 ○2 ○3 ○4 ○5 ○6	is common
is easy to understand	○1 ○2 ○3 ○4 ○5 ○6	is hard to understand
is rather comprehensive	○1 ○2 ○3 ○4 ○5 ○6	is rather limited

4. Additional Features at www.website.net

Unfortunately, www.website.net, does not possess adequate financial means to support the additional features. Therefore, it may become necessary to charge users a small fee for them.

In the following five questions, we would like you to indicate your willingness to pay a minimal sum for each of the additional features.

	Never	Unlikely	I am unsure	Possibly	Definitely
I would be willing to pay a minimal sum for bigger pictures that can be enhanced.	○	○	○	○	○
I would be willing to pay a minimal sum for hyperlinks to other useful websites concerning dermatology.	○	○	○	○	○
I would be willing to pay a minimal sum for information about alternative treatments.	○	○	○	○	○
I would be willing to pay a minimal sum for a search function that would deliver faster and more precise information	○	○	○	○	○
I would be willing to pay a minimal sum for access to terminological support for medical terms.	○	○	○	○	○

4. Additional Features at www.website.net

If you were asked to choose 3 from the 5 considered additional features, which ones would be the most interesting for you? Please indicate this to us by clicking in front of each of the 3 features that you would most prefer.

- ☐ Bigger and enhanced pictures
- ☐ Hyperlinks to other useful dermatological websites
- ☐ Information on alternative treatments for dermatological diseases
- ☐ Search function that makes your search target easier to find.
- ☐ Access to more consumer-friendly language (not so many medical terms).

Suppose you were to pay for the additional features that were discussed above with a flat rate fee. What time frame would you prefer this flat rate free to be valid for? Please indicate the time frame that best suits you clicking on your preferred choice.

- ○ A daily flat rate payment.
- ○ A weekly flat rate payment.
- ○ A monthly flat rate payment.
- ◉ A yearly flat rate payment.

5. Demographic Data

Please answer the following questions by selecting your answer from the scroll down list.

Your country of origin:	Albania
Your country of residence:	Albania
How high are the tuition fees for higher education in your country of origin?	Very high
Do you hold a position in medical care?	No
How good do you think the national health care system in your country of origin is?	Very good

A-3.5 Questionnaire (Experiment): Product Valuation (English Version)

(relevant sections)

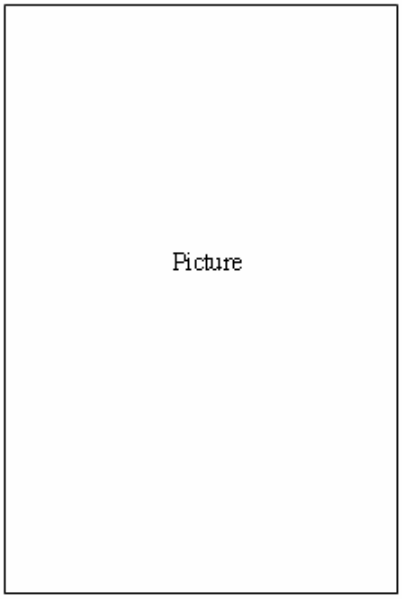
Registration - Step 1 of 3

Please fill out the following registration form in order to participate.

1. Type in a username for this session.	<?php echo $username?>
2. Email-address:	<?php echo $f2?>
3. Country of residence:	
4. City of residence:	<?php echo $f4?>
5. Country of origin:	
6. Native language:	<?php echo $f6?>
7. How many years did you live in your native country?	<?php e
8. Your knowledge of English:	Please Choose
9. Your age:	<?php e
10. Your gender:	>Male >Female
11. How do you finance your education?	
12. Your monthly allowance:	
13. Your computer facilities at home:	
14. Do you pay for any services in the Internet?* (*Services are: email-services, cellphone services, ring tones, games, software download, webspace, consultancy/advisory services, adult content pages) We don't consider as a service: purchasing books, cds, records, purchasing merchandise via ebay.)	>No >Yes
15. How much did you pay approximately in 2004 for Internet services?	<?php >€

Get to know Website A

As of now, imagine you or someone else has the disease you see on the photo.



Please go to www.websiteA.edu to find out more information about your disease. You have 10 minutes to use this website in order to answer the following two questions.

What disease could he possibly have?

What is the cause of this disease?

We assure you that this is enough time to search for the needed information. After 10 minutes we will ask you to continue with the next part of this experiment. If you need less than 10 minutes to answers the questions concerning the picture, don't hesitate to continue with the experiment.

When you have answered the question or when the 10min are over, please click here

Reasons for visiting Website A

For what reasons could you imagine a person would use www.websiteA.edu? Please check all that apply.

- ☐ Time restrictions (i.e. no time to go to the physician.)
- ☐ Cost restrictions (i.e. It is too expensive to visit a physician.)
- ☐ Need for alternative information
- ☐ Embarrassment
- ☐ Neglibility of disease (i.e. It's not worth seeing a physician.)
- ☐ General information needs
- ☐ Other, please specify:

Introduction new features

The E-Health-Portal is planning on adding new features to its website. We would like to find out which features you find most interesting. Please click here to view them.

Reasons for money allocation

Why did you choose the money allocation that you did? Please specify to which extent you agree with the following statements. Click on the corresponding buttons

	I strongly agree					I don't agree at all
... because complete information is important for me.	○	○	○	○	○	○
... because fast access to needed information is important for me.	○	○	○	○	○	○
... because objective information is important for me.	○	○	○	○	○	○
... because it is important for me that the information is easy to understand.	○	○	○	○	○	○
... because credible information is important for me.	○	○	○	○	○	○
... because an efficient use of the website is important for me.	○	○	○	○	○	○
... because websites that are easy to use are important for me.	○	○	○	○	○	○
... because comprehensible explanations are important for me.	○	○	○	○	○	○

For which Internet Services do you pay?

Please click on all that apply.

- ☐ Software downloads
- ☐ Webspace (email, website)
- ☐ General information (newspaper, articles, etc.)
- ☐ Customized information (e.g. job search, catalogues)
- ☐ Games, ringtones, pictures, etc.
- ☐ Specific services for private use (e.g. online advisory services, online banking, etc.)
- ☐ Specific services for scientific use (e.g. databases)
- ☐ Specific services for business use (e.g. ASP)

Why don't you pay for any Internet Services?

Please click on all that apply.

- ☐ too expensive for the money that I am able to spend every month
- ☐ unsatisfactory relationship price – service/quality
- ☐ I don't need them.
- ☐ Every paid service can be found also for free on the Internet.
- ☐ I don't have access to a credit card.
- ☐ I am afraid personal information will be exposed and/or abused (e.g. credit card information)
- ☐ The purchase process is too annoying.
- ☐ I don't know any paid service.

A-3.6 Selected Statistics: Product Valuation (Study 6)

Study A: Website Survey

(1) Factor Analysis: Measure 1

Rotated Component Matrix(a)

	Component			
	1	2	3	4
P3_1			,804	
P3_2		,637		
P3_3_rec				
P3_4		,680	,510	
P3_5	,480		,556	
P3_6	,598		,378	
P3_7	,812			
P3_8	,752			-,453
P3_9_rec	-,627			,312
P3_10	,789	,394		
P3_11	,455	,677		
P3_12_rec	-			,893
P3_13	,381	,659	,442	
P3_14		,765		
P3_15		,798		
P3_16	,698	,329	,334	
P3_17	,578	,517		

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 5 iterations.

(2) Health Care System: Measure 1

Ranks

	Quality of health care system in country of origin	N	Mean Rank
The information I received from the website:uses too few/many medical terms	very bad	7	29,50
	Bad	8	41,63
	Satisfactory	23	40,33
	Good	27	34,06
	Very good	10	46,35
	Total	75	
The information I received from the website: is easy/hard to find	very bad	7	37,64
	Bad	6	35,67
	Satisfactory	23	38,65
	Good	24	30,85
	Very good	10	37,80
	Total	70	
The information I received from the website: is easy/hard to access	very bad	7	44,36
	Bad	6	41,50
	Satisfactory	23	36,26
	Good	24	30,63
	Very good	10	35,65
	Total	70	
The information I received from the website:is easy/hard to understand	very bad	7	44,93
	Bad	7	40,43
	Satisfactory	23	34,83
	Good	24	31,88
	Very good	10	39,25
	Total	71	

Test Statistics(a,b)

	The information I received from the website:uses too few/many medical terms	*The information I received from the website: is easy/hard to find*	*The information I received from the website: is easy/hard to access*	*The information I received from the website:is easy/hard to understand*
Chi-Square	4,355	2,144	3,472	3,103
Df	4	4	4	4
Asymp. Sig.	,360	,709	,482	,541

a Kruskal Wallis Test

b Grouping Variable: Quality of health care system in country of origin

(3) Health Care System: Measure 2

Ranks

	Quality of health care system in country of origin	N	Mean Rank
Pictures	very bad	8	42,13
	Bad	9	38,83
	Satisfactory	23	39,98
	Good	28	43,54
	Very good	11	30,45
	Total	79	
Alternative Treatments	very bad	8	47,69
	Bad	9	44,94
	Satisfactory	23	45,33
	Good	28	31,46
	Very good	11	40,95
	Total	79	

Test Statistics(a,b)

	Pictures	*Alternative Treatments*
Chi-Square	4,191	8,764
Df	4	4
Asymp. Sig.	,381	,067

a Kruskal Wallis Test

b Grouping Variable: Quality of health care system in country of origin

(3) Health Care System – Measure 3

Ranks

	Quality of health care system in country of origin	N	Mean Rank
Pictures	very bad	8	45,81
	Bad	9	25,39
	Satisfactory	23	36,28
	Good	27	44,44
	Very good	10	37,35
	Total	77	
Hyperlinks	very bad	8	49,31
	Bad	9	41,06
	Satisfactory	23	34,61
	Good	27	37,93
	Very good	10	41,90
	Total	77	
Alternative Treatments	very bad	8	46,44
	Bad	9	35,33
	Satisfactory	22	37,05
	Good	27	35,74
	Very good	10	45,65
	Total	76	
Search function	very bad	8	44,31
	Bad	9	34,94
	Satisfactory	23	37,80
	Good	27	37,74
	Very good	10	44,55
	Total	77	

Test Statistics(a,b)

	Pictures	*Hyperlinks*	*Alternative Treatments*	*Search function*
Chi-Square	6,768	3,231	3,172	1,642
Df	4	4	4	4
Asymp. Sig.	,149	,520	,529	,801

a Kruskal Wallis Test

b Grouping Variable: Quality of health care system in country of origin

Study B: Experiment

(1) Factor Analysis: Remunerated Online Services

Rotated Component Matrix(a)

	Component	
	1	2
money for software downloads	,675	,468
money for webspace	,689	,353
money for general information	,840	
money for customized information	,860	
money for games, ringtones, pictures	,675	
money for specific services for private use	,728	,376
money for specific services for scientific use		,982
money for specific services for business use		,982

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 3 iterations.

Rotated Component Matrix(a)

	Component		
	1	2	3
no money for Internet Service because too expensive	,536	,494	
no money for Internet Services because inconvenient relationship price – quality	,425		-,646
no money for Internet Services because not needed.	,419		,691
no money for Internet Services because also free	,741		
no money for Internet Services because no credit card		,729	
no money for Internet Services because information will be abused	,762		
no money for Internet Services because purchase process annoying			,576
no money for Internet Services because not known.		,712	

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 5 iterations.

Results lower than 0.3 were suppressed.

(2) Correlations between Independent Variables and Control Variables

Group Statistics

	2 Gruppen UA	N	Mean	Std. Deviation	Std. Error Mean
money_private	low (<50)	3	-,3066150	,00000000	,00000000
	high (> 50)	96	-,2250241	,29025818	,02962435
money_professional	low (<50)	3	-,1335528	,00000000	,00000000
	high (> 50)	96	-,0832629	,18752069	,01913875
nomoney_moneyandprivacy	low (<50)	3	-,3206680	1,10157818	,63599646
	high (> 50)	96	,3566906	1,01725013	,10382266
nomoney_nofacilities	low (<50)	3	-,3024178	,12718542	,07343053
	high (> 50)	96	,0754844	,88456210	,09028024
nomoney_noneed	low (<50)	3	-,4477046	,09962815	,05752034
	high (> 50)	96	,1787874	1,10380218	,11265634

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
money_private	Equal variances assumed	1,078	,302	-,484	97	,629	-,08159089	,16841539	-,41584881	,25266703
	Equal variances not assumed			-2,754	95,000	,007	-,08159089	,02962435	-,14040266	-,02277911
money_professional	Equal variances assumed	,950	,332	-,462	97	,645	-,05028990	,10880441	-,26623653	,16565673
	Equal variances not assumed			-2,628	95,000	,010	-,05028990	,01913875	-,08828512	-,01229467
nomoney_moneyandprivacy	Equal variances assumed	,002	,962	-1,134	97	,260	-,67735868	,59747649	-1,86318412	,50846675
	Equal variances not assumed			-1,051	2,108	,399	-,67735868	,64441496	-3,31829312	1,96357575
nomoney_nofacilities	Equal variances assumed	2,601	,110	-,736	97	,463	-,37790225	,51335773	-1,39677523	,64097074
	Equal variances not assumed			-3,247	12,037	,007	-,37790225	,11637253	-,63136960	-,12443490
nomoney_noneed	Equal variances assumed	5,364	,023	-,978	97	,330	-,62649208	,64050980	-1,89772672	,64474257
	Equal variances not assumed			-4,953	35,710	,000	-,62649208	,12649126	-,88310060	-,36988356

Group Statistics

	2 Gruppen LTO	N	Mean	Std. Deviation	Std. Error Mean
money_private	low (<50)	31	-,0384625	,84499253	,15176514
	high (> 50)	30	,3281155	1,66708355	,30436642
money_professional	low (<50)	31	-,0475590	,23280128	,04181234
	high (> 50)	30	-,2171592	,19012098	,03471118
nomoney_moneyandprivacy	low (<50)	31	,0549276	1,10683474	,19879339
	high (> 50)	30	-,3959381	,79588959	,14530889
nomoney_nofacilities	low (<50)	31	,1811932	,95645119	,17178371
	high (> 50)	30	-,1921793	,52933672	,09664322
nomoney_noneed	low (<50)	31	,0615762	1,00218493	,17999773
	high (> 50)	30	-,1680821	1,23973173	,22634301

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
money_private	Equal variances assumed	4,443	,039	-1,089	59	,281	-,36657794	,33676897	-1,04045110	,30729522
	Equal variances not assumed			-1,078	42,664	,287	-,36657794	,34010524	-1,05262191	,31946603
money_professional	Equal variances assumed	,066	,798	3,111	59	,003	,16960018	,05452409	,06049773	,27870263
	Equal variances not assumed			3,121	57,398	,003	,16960018	,05434279	,06079697	,27840340
nomoney_moneyandprivacy	Equal variances assumed	4,210	,045	1,821	59	,074	,45086572	,24754887	-,04447842	,94620987
	Equal variances not assumed			1,831	54,521	,073	,45086572	,24623868	-,04270509	,94443653
nomoney_nofacilities	Equal variances assumed	7,428	,008	1,878	59	,065	,37337254	,19885544	-,02453628	,77128136
	Equal variances not assumed			1,894	47,113	,064	,37337254	,19710290	-,02312221	,76986729
nomoney_noneed	Equal variances assumed	,228	,635	,797	59	,429	,22965830	,28818048	-,34698951	,80630611
	Equal variances not assumed			,794	55,732	,430	,22965830	,28918911	-,34971865	,80903526

(3) Factor Analysis – Measure 1

Rotated Component Matrix(a)

	Component	
	1	2
time restrictions as reason	,631	
cost restrictions as reason	,792	
need for alternative information as reason	,	,781
embarrassment as reason	,476	,473
neglibility of disease as reason	,653	
general information needs as reason		,799

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 3 iterations.

Results lower than 0.3 were suppressed.

(4) Correlations between Money Allocation and Reasons for Money Allocation

Correlations

		money allocation for hyperlinks	*money allocation for alternative information*	*money allocation for search filter*	*money allocation for terminological support*
complete information as reason for money allocation	Pearson Correlation	,014	,019	,254(*)	,153
	Sig. (2-tailed)	,897	,868	,021	,167
	N	83	83	83	83
fast information access as reason for money allocation	Pearson Correlation	,071	,254(*)	,076	,132
	Sig. (2-tailed)	,522	,021	,494	,234
	N	83	83	83	83
objective information as reason for money allocation	Pearson Correlation	-,007	,095	,279(*)	,118
	Sig. (2-tailed)	,954	,394	,011	,291
	N	82	82	82	82
easy understanding as reason for money allocation	Pearson Correlation	,104	,259(*)	,198	,091
	Sig. (2-tailed)	,350	,018	,073	,416
	N	83	83	83	83
credible information as reason for money allocation	Pearson Correlation	-,031	,125	,154	,193
	Sig. (2-tailed)	,779	,262	,168	,083
	N	82	82	82	82
efficient use as reason for money allocation	Pearson Correlation	,135	,253(*)	,203	,152
	Sig. (2-tailed)	,231	,023	,069	,175
	N	81	81	81	81
easy websites as reason for money allocation	Pearson Correlation	,098	,220(*)	,138	,071
	Sig. (2-tailed)	,376	,046	,214	,524
	N	83	83	83	83
comprehensible explanations as reason for money allocation	Pearson Correlation	,208	,137	,157	,011
	Sig. (2-tailed)	,061	,220	,159	,920
	N	82	82	82	82

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

A-3.7 Questionnaire: Attitude towards Data Disclosure

(relevant sections)

1. Questions concerning Online Registration

progress

In this section, we are interested in your general experience with filling out online registration forms at websites.
Please indicate your agreement/disagreement with each of the following statements.

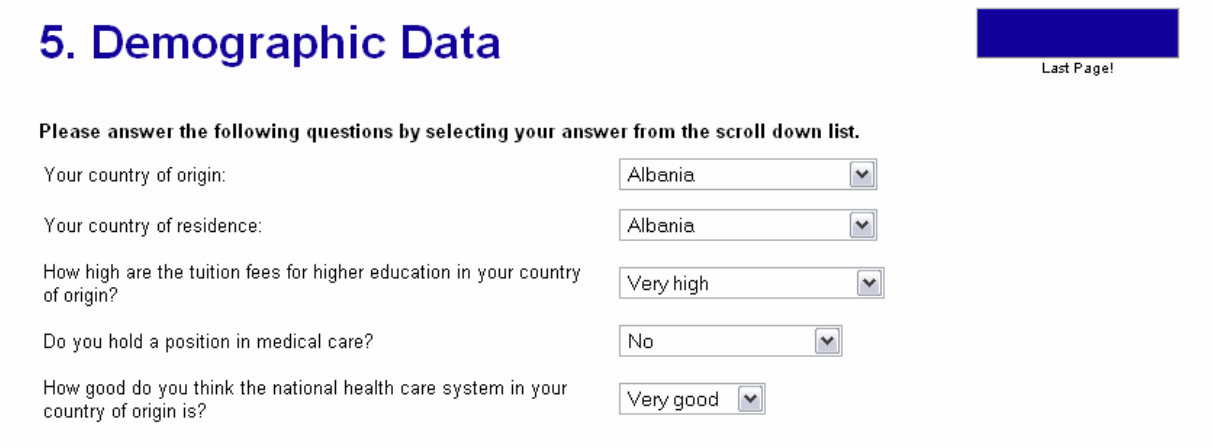
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
Online registration takes too much time.	○	○	○	○	◉
I would be willing to register online in exchange for personalized/customized services.	○	○	○	○	○
I see online registration as an invasion of my privacy.	○	○	○	○	○
Registering online is the price I pay in exchange for having access to information on a website.	○	○	○	○	○
I would refrain from registering online because I cannot be sure how my data is going to be used.	○	○	○	○	○
I value being able to visit websites in an anonymous manner.	○	○	○	○	○

4. Additional Features at www.website.net

progress

www.website.net is considering requiring its users to register online for the purpose of delivering some additional service features.
Please indicate how comfortable you would feel to provide each of the following types of information to the www.website.net web site.

	I would never provide this information.	Uncomfortable	Doubtful	Comfortable	Completely Comfortable
Your country of origin	○	○	○	○	○
Your medical knowledge	○	○	○	○	○
Your specific medical profession	○	○	○	○	○
Your e-mail address	○	○	○	○	○
Brief information on your current health or medical history.	○	○	○	○	○



A–3.8 Selected Statistics: Impact of Culture on Data Disclosure (Study 7)

(1) Factor Analysis: Attitudes towards Data Disclosure

Rotated Component Matrix(a)

	Component	
	1	2
P1_1_rec	,702	
P1_2		,717
P1_3_rec	,653	
P1_4		,841
P1_6_rec	,673	
P1_5_rec	,740	

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 3 iterations.

(2) Insignificant Impact of Power Distant and Individualism on Attitude towards Data Disclosure if Regarded Separately.

Multivariate Tests(b)

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	,010	,472(a)	2,000	97,000	,625
	Wilks' Lambda	,990	,472(a)	2,000	97,000	,625
	Hotelling's Trace	,010	,472(a)	2,000	97,000	,625
	Roy's Largest Root	,010	,472(a)	2,000	97,000	,625
PD_group	Pillai's Trace	,008	,402(a)	2,000	97,000	,670
	Wilks' Lambda	,992	,402(a)	2,000	97,000	,670
	Hotelling's Trace	,008	,402(a)	2,000	97,000	,670
	Roy's Largest Root	,008	,402(a)	2,000	97,000	,670
IND_group	Pillai's Trace	,010	,493(a)	2,000	97,000	,612
	Wilks' Lambda	,990	,493(a)	2,000	97,000	,612
	Hotelling's Trace	,010	,493(a)	2,000	97,000	,612
	Roy's Largest Root	,010	,493(a)	2,000	97,000	,612
PD_group * IND_group	Pillai's Trace	,005	,265(a)	2,000	97,000	,768
	Wilks' Lambda	,995	,265(a)	2,000	97,000	,768
	Hotelling's Trace	,005	,265(a)	2,000	97,000	,768
	Roy's Largest Root	,005	,265(a)	2,000	97,000	,768

a Exact statistic

b Design: Intercept+PD_group+IND_group+PD_group * IND_group

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Privacy_Contra	3,480(a)	3	1,160	1,109	,349
	Privacy_Pro	2,789(b)	3	,930	,894	,447
Intercept	Privacy_Contra	,021	1	,021	,020	,887
	Privacy_Pro	,975	1	,975	,938	,335
PD_group	Privacy_Contra	,707	1	,707	,676	,413
	Privacy_Pro	,154	1	,154	,148	,701
IND_group	Privacy_Contra	,019	1	,019	,019	,892
	Privacy_Pro	1,021	1	1,021	,982	,324
PD_group * IND_group	Privacy_Contra	,010	1	,010	,010	,921
	Privacy_Pro	,548	1	,548	,527	,470
Error	Privacy_Contra	102,487	98	1,046		
	Privacy_Pro	101,905	98	1,040		
Total	Privacy_Contra	106,363	102			
	Privacy_Pro	104,715	102			
Corrected Total	Privacy_Contra	105,967	101			
	Privacy_Pro	104,694	101			

a R Squared = ,033 (Adjusted R Squared = ,003)

b R Squared = ,027 (Adjusted R Squared = -,003)

(3) (Almost) Significant Impact of (Power Distance and Individualism) on Attitude towards Data Disclosure

Correlations

		Zscore(Cult_group)	Privacy_Contra	Privacy_Pro
Zscore(Cult_group)	Pearson Correlation	1	,180	,170
	Sig. (2-tailed)		,053	,069
	N	131	116	116

Regression

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	Cult_group(a)	.	Enter

a All requested variables entered.

b Dependent Variable: Privacy_Contra

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,180(a)	,032	,024	1,02716506

a Predictors: (Constant), Cult_group

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4,038	1	4,038	3,827	,053(a)
	Residual	120,278	114	1,055		
	Total	124,316	115			

a Predictors: (Constant), Cult_group

b Dependent Variable: Privacy_Contra

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-,574	,294		-1,956	,053
	Cult_group	,375	,192	,180	1,956	,053

a Dependent Variable: Privacy_Contra

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	Cult_group(a)	.	Enter

a All requested variables entered.

b Dependent Variable: Privacy_Pro

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,170(a)	,029	,020	,96418747

a Predictors: (Constant), Cult_group

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3,141	1	3,141	3,379	,069(a)
	Residual	105,981	114	,930		
	Total	109,122	115			

a Predictors: (Constant), Cult_group

b Dependent Variable: Privacy_Pro

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-,430	,276		-1,561	,121
	Cult_group	,331	,180	,170	1,838	,069

a Dependent Variable: Privacy_Pro

(4) Factor Analysis: Willingness to disclose data

Rotated Component Matrix(a)

	Component	
	1	2
provide information about country of origin	,729	
provide information about medical knowledge	,945	
provide specific information about medical profession	,950	
Provide e-mail address	,408	,700
provide brief information on current health or medical history		,917

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 3 iterations.

(5) Insignificant Impact of Culture on Willingness to disclose data

Multivariate Tests(b)

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	,004	,182(a)	2,000	91,000	,834
	Wilks' Lambda	,996	,182(a)	2,000	91,000	,834
	Hotelling's Trace	,004	,182(a)	2,000	91,000	,834
	Roy's Largest Root	,004	,182(a)	2,000	91,000	,834
Cult_group	Pillai's Trace	,024	1,112(a)	2,000	91,000	,333
	Wilks' Lambda	,976	1,112(a)	2,000	91,000	,333
	Hotelling's Trace	,024	1,112(a)	2,000	91,000	,333
	Roy's Largest Root	,024	1,112(a)	2,000	91,000	,333

a Exact statistic

b Design: Intercept+Cult_group

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	data_privacyinsensitive	,372(a)	1	,372	,370	,545
	data_privacysensitive	1,846(b)	1	1,846	1,863	,176
Intercept	data_privacyinsensitive	,061	1	,061	,060	,806
	data_privacysensitive	,302	1	,302	,304	,582
Cult_group	data_privacyinsensitive	,372	1	,372	,370	,545
	data_privacysensitive	1,846	1	1,846	1,863	,176
Error	data_privacyinsensitive	92,628	92	1,007		
	data_privacysensitive	91,154	92	,991		
Total	data_privacyinsensitive	93,000	94			
	data_privacysensitive	93,000	94			
Corrected Total	data_privacyinsensitive	93,000	93			
	data_privacysensitive	93,000	93			

a R Squared = ,004 (Adjusted R Squared = -,007)

b R Squared = ,020 (Adjusted R Squared = ,009)

(6) Significant Impact of Attitude towards Disclosing Data on Willingness to disclose data

Multivariate Tests(b)

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	,028	1,198(a)	2,000	84,000	,307
	Wilks' Lambda	,972	1,198(a)	2,000	84,000	,307
	Hotelling's Trace	,029	1,198(a)	2,000	84,000	,307
	Roy's Largest Root	,029	1,198(a)	2,000	84,000	,307
FAC1_attitude	Pillai's Trace	,103	4,845(a)	2,000	84,000	,010
	Wilks' Lambda	,897	4,845(a)	2,000	84,000	,010
	Hotelling's Trace	,115	4,845(a)	2,000	84,000	,010
	Roy's Largest Root	,115	4,845(a)	2,000	84,000	,010
FAC2_attitude	Pillai's Trace	,057	2,536(a)	2,000	84,000	,085
	Wilks' Lambda	,943	2,536(a)	2,000	84,000	,085
	Hotelling's Trace	,060	2,536(a)	2,000	84,000	,085
	Roy's Largest Root	,060	2,536(a)	2,000	84,000	,085

a Exact statistic

b Design: Intercept+FAC1_attitude+FAC2_attitude

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	data_privacyinsensitive	4,549(a)	2	2,275	3,535	,034
	data_privacysensitive	4,387(b)	2	2,194	2,247	,112
Intercept	data_privacyinsensitive	1,055	1	1,055	1,640	,204
	data_privacysensitive	,313	1	,313	,320	,573
FAC1_attitude	data_privacyinsensitive	3,374	1	3,374	5,243	,025
	data_privacysensitive	2,346	1	2,346	2,403	,125
FAC2_attitude	data_privacyinsensitive	1,188	1	1,188	1,845	,178
	data_privacysensitive	2,055	1	2,055	2,106	,150
Error	data_privacyinsensitive	54,702	85	,644		
	data_privacysensitive	82,973	85	,976		
Total	data_privacyinsensitive	60,022	88			
	data_privacysensitive	87,543	88			
Corrected Total	data_privacyinsensitive	59,251	87			
	data_privacysensitive	87,360	87			

a R Squared = ,077 (Adjusted R Squared = ,055)

b R Squared = ,050 (Adjusted R Squared = ,028)

Chapter 4

A-4.1 Number of Website Visitors, Internet Users, Percentage of Webhosts and Derived Statistics ordered by Language (Study 8)

Source of Internet Statistics: www.glreach.com and Languages and Internet

February 2005

Language	*Numbers of Website Visitors*	*Numbers of Internet Users (in Millions)*	*Percentage of Web-Hosts*	*Percentage of Webhosts/ 1000 Internet Users*	*Website Visitors/ Internet Users (in 1/1000 %)*
English	57956	295,4	68,40%	231,55%	19,62%
German	17362	55,3	5,80%	104,88%	31,40%
Spanish	7489	72	2,40%	33,33%	10,40%
French	5269	33,3	3%	90,09%	15,82%
Portuguese	5269	24,4	1,40%	57,38%	21,59%
Russian	519	6,5	1,90%	292,31%	7,98%
Chinese	456	110	3,90%	35,45%	0,41%
Japanese	302	67,1	5,90%	87,93%	0,45%

March 2005

Language	*Numbers of Website Visitors*	*Numbers of Internet Users (in Millions)*	*Percentage of Web-Hosts*	*Percentage of Webhosts/ 1000 Internet Users*	*Website Visitors/ Internet Users (in 1/1000 %)*
English	68030	295,4	68,40%	231,55%	23,03%
German	18619	55,3	5,80%	104,88%	33,67%
Spanish	8298	72	2,40%	33,33%	11,53%
Portuguese	7280	24,4	3%	90,09%	29,84%
French	5870	33,3	1,40%	57,38%	17,63%
Chinese	759	110	1,90%	292,31%	0,69%
Russian	588	6,5	3,90%	35,45%	9,05%
Japanese	266	67,1	5,90%	87,93%	0,40%

April 2005

Language	*Numbers of*	*Numbers of Internet*	*Percentage of Web-*	*Percentage of Webhosts/ 1000*	*Website Visitors/ Internet*

	Website Visitors	*Users (in Millions)*	*Hosts*	*Internet Users*	*Users (in 1/1000 %)*
English	42944	295,4	68,40%	231,55%	14,54%
German	11054	55,3	5,80%	104,88%	19,99%
Spanish	5615	72	2,40%	33,33%	7,80%
Portuguese	4747	24,4	3%	90,09%	19,45%
French	3642	33,3	1,40%	57,38%	10,94%
Chinese	398	110	1,90%	292,31%	0,36%
Russian	337	6,5	3,90%	35,45%	5,18%
Japanese	303	67,1	5,90%	87,93%	0,45%

A-4.2 Questionnaire (Website A): Impact of Language on Website Satisfaction in the Context of the World Wide Web

1a. Language Use

Progress

The following questions concern your linguistic habits.

Your mother tongue? (Please type in.)

In which language do you read this website? --Please select--

	What is your proficiency level with the language you use for reading this website?	**What is your proficiency level in English?** (Please answer this question even if your native tongue is English.)
Native speakers		
My native tongue	○	○
Non-native speakers		
Fluently, mother tongue-like	○	○
Communication and comprehension are possible without any problems in almost any situation.	○	○
Depending on the situation, communication and comprehensions are sometimes easy, sometimes difficult.	○	○
Communication and comprehension are possible only with lots of effort.	○	○
My knowledge of the language is rudimentary. Communication is hardly possible.	○	○

1b. Language Use

Progress

On a scale from 1 (I completely agree.) to 7 (I completely disagree.), to which extent do you agree with the following statements?

	I completely agree.						I completely disagree.
	1	2	3	4	5	6	7
There are enough websites in my native tongue on the Internet to allow me to find information about almost any topic.	○	○	○	○	○	○	○
If this website was offered in my mother tongue, I would probably check whether more or different information is offered in other language versions.	○	○	○	○	○	○	○
Medical information can be rarely found in my native tongue on the Internet.	○	○	○	○	○	○	○
If this website was offered in my native tongue, I would probably save a lot of time while surfing.	○	○	○	○	○	○	○
I understand that this website is not offered in my native tongue.	○	○	○	○	○	○	○

1b. Language Use

On a scale from 1 (I completely agree.) to 7 (I completely disagree.), to which extent do you agree with the following statements?

	I completely agree.						I completely disagree.
	1	2	3	4	5	6	7
There are enough websites on the Internet in my native language that I can find information on almost any topic.	○	○	○	○	○	○	○
On a multilingual website I usually check whether more or different information is offered in other language versions.	○	○	○	○	○	○	○
On the Internet, medical information is offered only rarely in my native tongue.	○	○	○	○	○	○	○
If this website was not offered in my native tongue, I would probably need more time for surfing.	○	○	○	○	○	○	○
I would not understand why, if this website was not offered in my tongue.	○	○	○	○	○	○	○

1c. Language Use

On a scale from 1 (I completely agree.) to 7 (I completely disagree.), to which extent do you agree with the following statements?

	I completely agree.						I completely disagree.
	1	2	3	4	5	6	7
If this website was offered in my native tongue I would probably surf much more here.	○	○	○	○	○	○	○
If this website was offered in my native tongue I would probably need less time for finding information. .	○	○	○	○	○	○	○
I find it unusual that this website is not offered in my native tongue.	○	○	○	○	○	○	○
If this website was offered in my native tongue, it would probably be less demanding to find information.	○	○	○	○	○	○	○

1c. Language Use

On a scale from 1 (I completely agree.) to 7 (I completely disagree.), to which extent do you agree with the following statements?

	I completely agree.						I completely disagree.
	1	2	3	4	5	6	7
If this website was not offered in my native tongue, I would probably surf less here.	○	○	○	○	○	○	○
I find it unusual that this website is offered in my native tongue.	○	○	○	○	○	○	○
If this website was not offered in my native tongue, it would probably be much more demanding to find information.	○	○	○	○	○	○	○
It would probably take me more time to find information, if this website was not offered in my native tongue.	○	○	○	○	○	○	○

Please rate **the result achieved through your visit** to this website by the means of the following adjectives. Indicate, for each pair, which adjective best describes the result of your visit.

You can differentiate your judgement on a scale from 1 to 7.

	1	2	3	4	5	6	7	
unsatisfied	○	○	○	○	○	○	○	satisfied
pleased	○	○	○	○	○	○	○	disappointed
positive	○	○	○	○	○	○	○	negative
discontented	○	○	○	○	○	○	○	contented

4. Demographic Data

Finally, we would like to ask you some general questions.
Please answer the following questions by clicking on the applicable answer from the list.

Where did you grow up?	--Please select.--
In which country do you presently reside?	--Please select.--
Your age:	(Please type in.)
Your gender:	male female
Do you work in the field of medicine?	--Please select.--
How many years have you been using the Internet?	for --Please select.--
How often have you used the Internet last week?	--Please select.--
Why do you visit this website?	--Please select.--

A-4.3 Questionnaire (Website B): Impact of Language on Website Satisfaction in the Context of the World Wide Web

1. Demographic Data

Progress

First, we would like to ask you some general questions.
Please answer the following questions by clicking on the applicable answer from the list.

Where did you grow up? --Please select--

In which country do you presently reside? --Please select--

Your native language: (Please type in.)

Your age: (Please type in.)

Your gender: male female

Which medical job do have? --Please select--

How many years have you been using the Internet? --Please select--

How often have you used the services on Telederm last month? --Please select--

Please rate on the following scale your English language proficiency level.

My mother tongue is English.

(Please click on "NEXT" at the end of this page.)

	Strongly agree 1	2	3	4	5	Strongly disagree 6
I find it easy to follow case discussions in English.						
I find it easy to ask questions and reply to questions in English.						

2a. Information exchange

The Discussion Forum

NOTE: All the questions refer exclusively to cases that are discussed in ENGLISH!

Please indicate your agreement/disagreement with each of the following statements on the following scale.

☐ *I can't tell. I haven't used the discussion forum yet.*
(Please click on "NEXT" at the end of this page.)

	Strongly agree 1	2	3	4	5	Strongly disagree 6
Generally, I obtain information in the discussion forum that is very difficult to find elsewhere.	○	○	○	○	○	○
Using the discussion forum is an efficient way to obtain information about certain topics.	○	○	○	○	○	○
Explanations are generally easier to understand in the discussion forum than elsewhere (e.g. reference books).	○	○	○	○	○	○
Overall, I find the English language discussion forum useful for obtaining information.	○	○	○	○	○	○

2b. Information Exchange

The Private Consultation Service

NOTE: All the questions refer exclusively to cases that are discussed in ENGLISH!

Please indicate your agreement/disagreement with each of the following statements on the following scale.

☐ *I can't tell. I haven't used the private consultation service yet.*
(Please click on "NEXT" at the end of this page.)

	Strongly agree 1	2	3	4	5	Strongly disagree 6
Generally, in the English language private consultation service section I get information that is very difficult to find elsewhere.	○	○	○	○	○	○
Using the English language private consultation service is an efficient way to get information about certain topics.	○	○	○	○	○	○
Answers to questions are easier to understand in the English language private consultation service than elsewhere (e.g. reference books).	○	○	○	○	○	○
Overall, I find the English language private consultation service useful for obtaining information.	○	○	○	○	○	○

I have also used the private consultation service in the following languages: ☐ German ☐ Italian ☐ Spanish ☐ Turkish ☐ Chinese ☐ Other

3a. Ease of Use

The Discussion Forum

NOTE: All of the questions refer exclusively to cases that are discussed in ENGLISH!

Please indicate your agreement/disagreement with each of the following statements on the following scale.

	Strongly agree 1	2	3	4	5	Strongly disagree 6
Generally it doesn't require much effort to follow and understand the English language discussions.	○	○	○	○	○	○
Generally it doesn't require much time to write replies and ask questions in English.	○	○	○	○	○	○
Posting questions in the discussion forum is an easy way to obtain information about specific topics.	○	○	○	○	○	○
This discussion forum is a fast way to get information about certain topics.	○	○	○	○	○	○
Overall I find it easy to use the discussion forum.	○	○	○	○	○	○

4a. General Impression

The Discussion Forum

NOTE: All of the questions refer exclusively to cases that are discussed in ENGLISH!

Following are a number of adjectives pairs. Please indicate, on the scale between the adjectives, the degree to which one of the two adjectives describes best your opinion.

All things considered, using the DISCUSSION FORUM for obtaining information about dermatological diseases is . . .

bad	extremely	quite	slightly	slightly	quite	extremely	**good**
foolish	extremely	quite	slightly	slightly	quite	extremely	**wise**
unfavorable	extremely	quite	slightly	slightly	quite	extremely	**favorable**
harmful	extremely	quite	slightly	slightly	quite	extremely	**beneficial**
negative	extremely	quite	slightly	slightly	quite	extremely	**positive**

4b. General Impression

The Private Consultation Service

NOTE: All of the questions refer exclusively to cases that are discussed in ENGLISH!

Following are a number of adjectives pairs. Please indicate on the scale between the adjectives to degree one of the two adjectives describes best your opinion.

All things considered, using the PRIVATE CONSULTATION SERVICE for obtaining information about dermatological diseases is . . .

bad	○ extremely	○ quite	○ slightly	○ slightly	○ quite	○ extremely	**good**
foolish	○ extremely	○ quite	○ slightly	○ slightly	○ quite	○ extremely	**wise**
unfavorable	○ extremely	○ quite	○ slightly	○ slightly	○ quite	○ extremely	**favorable**
harmful	○ extremely	○ quite	○ slightly	○ slightly	○ quite	○ extremely	**beneficial**
negative	○ extremely	○ quite	○ slightly	○ slightly	○ quite	○ extremely	**positive**

6a. Language Preferences

The Discussion Forum

On a scale of 1 (I completely agree.) to 6 (I completely disagree.), to which extent do you agree with the following statements?

	I completely agree.					I completely disagree.
	1	2	3	4	5	6
There are enough websites or chat-forums in English on the Internet to find medical information about almost any topic.	○	○	○	○	○	○
If cases in the discussion forum were discussed in languages other than English, I would probably (also) participate in at least on of those other discussion forums.	○	○	○	○	○	○
Medical information can rarely be found in English on the Internet.	○	○	○	○	○	○
If discussion forum cases were not discussed in English, I would probably lose a lot of time for reading the information.	○	○	○	○	○	○
I would be negatively surprised if discussion forum cases were not discussed in English.	○	○	○	○	○	○
If cases in the discussion forum were not discussed in English, I would probably use the discussion forum less often.	○	○	○	○	○	○
If cases in the discusssion forum were not discussed in English, it would probably take me more time to answer questions.	○	○	○	○	○	○
If cases in the discussion forum were not discussed in English, it would probably cost me more effort to participate.	○	○	○	○	○	○

6b. Language Preferences

The Private Consultation Service

On a scale of 1 (I completely agree.) to 6 (I completely disagree.), to which extent do you agree with the following statements?

	I completely agree.					I completely disagree.
	1	2	3	4	5	6
There are enough experts on the Consultation Services on the Internet that I can consult in my native tongue.	○	○	○	○	○	○
If experts offered consultation services in my native language, I would probably (also) send requests to experts with whom I have to communicate in English.	○	○	○	○	○	○
If I could consult experts in my native tongue , I would probably save a lot of time for reading the information.	○	○	○	○	○	○
I understand that I have to consult most experts with in English.	○	○	○	○	○	○
If I could consult experts in my native tongue, I would probably use the private consultation service more often.	○	○	○	○	○	○
If I could consult experts in my native tongue, it would probably take me less time to ask questions.	○	○	○	○	○	○
If I could consult experts in my native tongue, it would probably cost me less effort to use the service.	○	○	○	○	○	○

6b. Language Preferences

The Private Consultation Service

On a scale of 1 (I completely agree.) to 6 (I completely disagree.), to which extent do you agree with the following statements?

	I completely agree.					I completely disagree.
	1	2	3	4	5	6
There are enough experts for the Consultation Service that I can consult in English.	○	○	○	○	○	○
If experts offered consultation services in different languages, I would probably (also) send requests to experts with whom I have to communicate in a language other than English.	○	○	○	○	○	○
If there were no experts that I could consult with in English, I would probably lose a lot of time for reading the information.	○	○	○	○	○	○
I would be negatively surprised if I had to consult most experts in a language other than English.	○	○	○	○	○	○
If I couldn't consult experts in English, I would probably use the private consultation service less often.	○	○	○	○	○	○
If I couldn't consult an expert in my native tongue, it would probably take me more time to ask questions.	○	○	○	○	○	○
If I couldn't consult an expert in English, it would probably cost me more effort to use the service.	○	○	○	○	○	○

A-4.4 Selected Statistics: Impact of Language on Website Satisfaction in the Context of the World Wide Web (Study 9a)

(1) Factor Loadings for items composing Satisfaction (SAT), perceived amount of information online (ILO) and effort saved (SAVEFF).

Rotated Component Matrix(a)

	Component		
	1	2	3
SAT_1	,768		
SAT_2_rec	,905		
SAT_3_rec	,902		
SAT_4	,896		
ILO_1_rec			,834
ILO_3		-,321	,710
ILO_4_rec	-,359		,785
SAVEFF_1_rec		,822	
SAVEFF_2_rec		,877	
SAVEFF_3_rec		,851	
SAVEFF_4_rec		,892	

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 4 iterations.

Results smaller than 0.3 were suppressed.

(2) Test of normal distribution for constructs of Satisfaction (SAT), perceived amount of information online (ILO) and effort saved (SAVEFF).

One-Sample Kolmogorov-Smirnov Test

		SAVEFF_MEAN	ILO_MEAN	SAT_MEAN
N		25	28	29
Normal Parameters(a,b)	Mean	4,8300	4,5119	5,6121
	Std. Deviation	2,06499	1,81553	1,31710
Most Extreme Differences	Absolute	,234	,130	,162
	Positive	,178	,099	,146
	Negative	-,234	-,130	-,162
Kolmogorov-Smirnov Z		1,171	,690	,875
Asymp. Sig. (2-tailed)		,129	,727	,429

a Test distribution is Normal.

b Calculated from data.

(3) Regression Analysis testing for an impact of perceived amount of information online (ILO) and effort saved (SAVEFF) on Satisfaction (SAT),.

Regression

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	ILO_MEAN, SAVEFF_MEAN (a)	.	Enter

a All requested variables entered.

b Dependent Variable: SAT_MEAN

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,212(a)	,045	-,051	1,34989

a Predictors: (Constant), ILO_MEAN, SAVEFF_MEAN

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,708	2	,854	,469	,633(a)
	Residual	36,444	20	1,822		
	Total	38,152	22			

a Predictors: (Constant), ILO_MEAN, SAVEFF_MEAN

b Dependent Variable: SAT_MEAN

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6,510	1,189		5,476	,000
	SAVEFF_MEAN	-,038	,145	-,059	-,262	,796
	ILO_MEAN	-,165	,171	-,218	-,968	,345

a Dependent Variable: SAT_MEAN

(

3) Perceived amount of information online (ILO) per language.

Report

mother_tongue_num		ILO_MEAN
German	Mean	5,5556
	N	9
	Std. Deviation	1,28019
English	Mean	5,8889
	N	3
	Std. Deviation	1,01835
Spanish	Mean	2,5833
	N	4
	Std. Deviation	1,61876
French	Mean	3,0000
	N	1
	Std. Deviation	.
Portuguese	Mean	3,6667
	N	6
	Std. Deviation	1,57762
Total	Mean	4,4783
	N	23
	Std. Deviation	1,81416

(4) Average proficiency levels in English per language.

Report

mother_tongue_num		Proficiency Level in English
German	Mean	3,44
	N	9
	Std. Deviation	1,509
English	Mean	1,00
	N	3
	Std. Deviation	,000
Spanish	Mean	4,50
	N	4
	Std. Deviation	1,732
French	Mean	4,00
	N	1
	Std. Deviation	.
Portuguese	Mean	2,80
	N	5
	Std. Deviation	2,168
Total	Mean	3,18
	N	22
	Std. Deviation	1,816

A-4.5 Selected Statistics: Impact of Language on Website Satisfaction in the Context of the World Wide Web (Study 9b – Part I)

(1) Factor Loadings for items composing Satisfaction (SAT), perceived amount of information online (ILO) and effort saved (SAVEFF).

Rotated Component Matrix(a)

	Component		
	1	2	3
SAT_1	,916		
SAT_2	,868		
SAT_3	,925		
SAT_4	,938		
SAT_5	,908		
ILO_1			,846
ILO_2			,762
SAVEFF_1	,310	,870	
SAVEFF_2		,872	
SAVEFF_3		-,490	

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 4 iterations.

Results smaller than 0.3 were suppressed.

(2) Bivariate correlations between items composing Satisfaction (SAT) and effort saved (SAVEFF).

Correlations

			SAVEFF_1	*SAVEFF_2*	*SAVEFF_3*
Spearman's rho	SAT_1	Correlation Coefficient	,312(*)	,274(*)	
		Sig. (2-tailed)	,023	,045	
		N	53	54	
	SAT_2	Correlation Coefficient	,317(*)	,312(*)	
		Sig. (2-tailed)	,022	,023	
		N	52	53	
	SAT_3	Correlation Coefficient			
		Sig. (2-tailed)			
		N			
	SAT_4	Correlation Coefficient	,251		
		Sig. (2-tailed)	,072		
		N	52		
	SAT_5	Correlation Coefficient	,370(**)	,325(*)	
		Sig. (2-tailed)	,007	,018	
		N	52	53	

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

(3 Bivariate correlations between items composing effort saved (SAVEFF) and perceived amount of information online (ILO).

Correlations

			ILO_1	*ILO_2*	*ILO_3*
Spearman's rho	SAVEFF_1	Correlation Coefficient	-,181	,006	-,155
		Sig. (2-tailed)	,194	,964	,273
		N	53	52	52
	SAVEFF_2	Correlation Coefficient	-,064	,056	-,100
		Sig. (2-tailed)	,646	,690	,477
		N	54	53	53
	SAVEFF_3	Correlation Coefficient	-,146	,032	-,078
		Sig. (2-tailed)	,292	,820	,577
		N	54	53	53

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

A-4.6 Selected Statistics: Impact of Language on Website Satisfaction in the Context of the World Wide Web (Study 9b – Part II)

(1) Factor Loadings for items composing Satisfaction (SAT), perceived amount of information online (ILO) and effort saved (SAVEFF).

Rotated Component Matrix(a)

	Component	
	1	2
SAT_1	,958	
SAT_2	,956	
SAT_3	,962	
SAT_4	,966	
SAT_5	,938	
ILO_1		
ILO_2	,361	,405
SAVEFF_1		,916
SAVEFF_2		,940
SAVEFF_3		,900
SAVEFF_4		,944

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 3 iterations.

Results smaller than 0.3 were suppressed.

(2) Average proficiency levels in English per language.

Report

L1_num		I find it easy to follow case discussions in English.	I find it easy to ask questions and reply to questions in English.
Arabic	Mean	1,20	1,20
Bangla	Mean	2,00	2,00
Chinese	Mean	2,00	2,36
Czech	Mean	1,33	1,67
German	Mean	1,75	1,94
Dutch	Mean	1,50	2,00
Farsi/Persian	Mean	1,57	2,14
Finnish	Mean	1,00	1,00
French	Mean	2,00	3,50
Georgian	Mean	1,00	1,00
Greek	Mean	3,00	3,50
Gujarati	Mean	1,00	1,00
Hebrew	Mean	1,00	1,00
Hindi	Mean	1,50	1,50
Italian	Mean	2,26	2,53
Japanese	Mean	1,00	1,00
Latvia	Mean	1,00	1,00
Oriya	Mean	1,00	1,00
Portuguese	Mean	2,00	2,40
Punjabi	Mean	1,00	1,00
Romanian	Mean	3,00	4,00
Russia	Mean	1,50	1,50
Serbia	Mean	2,00	2,00
Slovenia	Mean	1,50	2,00
Spanish	Mean	1,33	1,67
Swedish	Mean	1,00	2,00
Tamil	Mean	1,00	1,00
Telugu	Mean	1,00	1,00
Turkish	Mean	1,94	2,18
Total	Mean	1,77	2,05

Empfangene Unterstützung und Hilfe durch Kollegen

- Prof. Dr. Berendt gab wertvolle Beiträge und Ideen für Kapitel 1.

- In Studie 3 sind Kommentare der Gutachter der Zeitschrift „New Review of Hypermedia and Multimedia“ eingeflossen.

- Steffan Baron und Gebhard Dettmold unterstützen mich bei der Überführung der Logfile-Daten in eine MySQL- und Oracle-Datenbank sowie bei der Nutzung von WUM und WUMprep.

- Nurfauza Jali und Matthias Fischmann programmierten die notwendigen Webseiten für Studie 4. Ideen für Studie 4 sind mit Professor Yeo diskutiert worden.

- Prof. Dr. Eisend lieferte die Grundlage für die Auswertung der Daten in Studie 5 und Studie 1. Ideen für Studie 5 und Studie 1 sind ebenfalls mit ihm diskutiert worden.

- Dr. Mandl programmierte den Webcrawler für Studie 8 und gab weitere Hinweise zur Entwicklung und Auswertung der Studie. Es flossen weiterhin Kommentare der Gutachter der Hawaii International Conference on Systems Sciences (HICSS-39) ein.

- Veit Köppen unterstütze mich in der Entwicklung von Studie 9.

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Whitehorse, d. 17.11.2005

Anett Kralisch

Eidesstattliche Erklärung

Hiermit erkläre ich, Anett Kralisch, dass ich mich bisher noch an keiner Institution einem Doktorexamen unterzogen habe. Ferner wurde die Dissertation bisher an noch keiner anderen Fakultät vorgelegt.

Whitehorse, d.17.11.2005

Anett Kralisch